Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
HUMAN RESOURCES RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *Human Resources Research Center*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

Studies of Distributed Practice: II. Learning and Retention of Paired-Adjective Lists with Two Levels of Intra-List Similarity: B. J. UNDERWOOD 153

The Effect of Motivation and Amount of Pre-Rest Practice upon Inhibitory Potential in Motor Learning: H. N. WASSERMAN 162

The Acquired Distinctiveness of Cues: The Role of Discriminative Verbal Responses in Facilitating the Acquisition of Discriminative Motor Responses: IRMA L. ROSSMAN AND A. E. GOSS 173

Anxiety and Strength of the UCS as Determiners of the Amount of Eyelid Conditioning: K. W. SPENCE AND JANET TAYLOR 183

Memory for Pleasant and Unpleasant Experiences: Some Methodological Considerations: R. H. TURNER AND J. A. BARLOW 189

Mediated Stimulus Generalization as a Factor in Sensory Pre-Conditioning: D. D. WICKENS AND G. E. BRIGGS 197

The Roles of Pattern and Apparent Distance in Determining the Color of Areas Seen Through Transparencies: W. C. H. PRENTICE, JOSEPHINE KRIMSKY, AND S. BARKER 201

The Ability to Make Absolute Judgments of Pitch: ALICE VAN KREVELEN 207

American Psychological Association

Vol. 42 No. 3 September 1951

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.00, or $14.00 annually. Single copies are $1.25. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., Prince and Lemon Sts., Lancaster, Pa., or 1515 Massachusetts Avenue, N.W., Washington 5, D. C.

This JOURNAL publishes articles which are intended to contribute toward the development of psychology as an experimental science. Experimental work with normal human subjects is favored over work with abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental psychology or for behavior theory. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, a single experimental study or an integrated series of studies accomplished simultaneously must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors who request them are supplied with 50 free reprints without covers. However, under the new APA reprint policy no free reprints will be supplied to authors submitting manuscripts after January 1, 1952. All of the cost of the author's alterations in galley proof is charged to the author, and one-half of the cost of tables, cuts, and formulas.

Immediate publication can be given to articles whose authors assume the full cost of publication. Such articles must undergo the usual review by the Consulting Editors of the JOURNAL. No free reprints are given in the case of these articles.

Address all articles submitted for regular publication or for immediate publication to the editor, Arthur W. Melton, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

The JOURNAL adheres to the conventions concerning bibliographies, citations, preparation of tables and cuts, etc., as described by J. E. Anderson and W. L. Valentine, The preparation of articles for publication in the journals of the American Psychological Association, *Psychol. Bull.*, 1944, 41, 345–376. Special attention is called to the style of numbers and abbreviations (p. 353). The term 'experimenter,' with its various forms, singular, plural, and possessive singular and plural, is abbreviated *E*, *E*s, *E*'s, and *E*s'; analogous abbreviations are used for the terms 'subject' (*S*) and 'observer' (*O*). All manuscripts must be typed double-spaced throughout, including tables, footnotes, references, and titles for figures. Tables, figures, footnotes, references, and titles for figures must be placed on separate pages at the end of the manuscript.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Journal of
Experimental Psychology

VOL. 42 NO. 3 SEPTEMBER, 1951

STUDIES OF DISTRIBUTED PRACTICE: II. LEARNING AND RETENTION OF PAIRED-ADJECTIVE LISTS WITH TWO LEVELS OF INTRA-LIST SIMILARITY

BY BENTON J. UNDERWOOD

Northwestern University[1]

In the first study in this series it was shown that serial lists of highly similar adjectives were learned more rapidly by distributed than by massed practice (**11**). Lists made up of dissimilar items tended to be learned more rapidly by distribution, but differences were not significant statistically. Thus, it was concluded that intra-task similarity is a variable which must be accounted for in any theory of distributed practice of verbal materials. Furthermore, the results demonstrated that the number of errors made during learning was directly related to facilitation produced by distributed practice. More errors per trial occurred under the distributed conditions than under the massed conditions. This fact was taken as evidence against the McGeoch version of the differential-forgetting theory, since that theory predicts a decrease in the number of errors following a rest period. On the positive side the evidence conformed to the general outline of Gibson's theory of verbal learning since that theory would predict an increase in error frequency following rest intervals, the increase being attributed to the recovery of generalization tendencies (**2**). Since Gibson's theory does not have a specific accounting of the influence of distributed practice on learning, details of the findings could not be checked against the theory.

In the present experiments the investigation has been extended to paired-associate lists with two degrees of similarity, one in which intra-list similarity is slight, and one in which it is very great. These lists are presented under different degrees of distribution. In addition, retention measurements have been taken after three different intervals.

PROCEDURE

Materials.—For each level of similarity three lists were required. A sample of the lists is presented below. In constructing the high-simi-

Low Similarity	*High Similarity*
lovely—resting	clumsy—twofold
uphill—filmy	double—soiled
oldish—instant	dirty—frigid
drastic—latent	arctic—devoid
pious—mature	awkward—unclean
rancid—barbed	vacant—dual
adrift—equal	paired—bungling
jellied—aware	frozen—empty
nimble—vagrant	filthy—icy
human—waxen	unfilled—graceless

[1] These experiments were made possible by a grant from The Graduate School, Northwestern University.

larity lists an attempt was made to maximize the similarity relationships among the items; thus, five sets of four highly synonymous words, taken from Haagen's lists (4), were employed. Similar items appeared both as stimuli and as responses; therefore, it is presumed that stimulus generalization, response generalization, and stimulus-response generalization occurred.

The lists of adjectives were presented by a Hull-type drum with a 2-sec. exposure of the stimulus followed by 2 sec. in which the stimulus and response were presented together. The lists were presented in three different orders to minimize serial learning. Anticipation learning was used throughout.

Conditions.—The common feature of the six experiments was the three degrees of distribution. In the most massed condition 4 sec. elapsed between trials; in a second condition, 30 sec. occurred between trials, and in the third condition, 2 min. was used as the distribution interval. The differences among the experiments are shown below. Exp. I, II, and III employed the low-similarity lists; Exp. IV, V, and VI, the high-similarity lists. The retention intervals were 10 min., 24 hr., and 48 hr.

Low Similarity	*Retention Interval*	*High Similarity*	*Retention Interval*
Exp. I	10 min.	Exp. IV	10 min.
Exp. II	24 hr.	Exp. V	24 hr.
Exp. III	48 hr.	Exp. VI	48 hr.

For all conditions the learning before the retention interval was carried to 8 out of 10 correct responses on a single trial. After the retention interval, relearning was carried to one errorless trial by massed practice.

For each experiment, 18 *Ss* were used. The three conditions of distribution were completely counterbalanced among these 18 *Ss*, with lists held constant from day to day. Since the original learning was under identical conditions for Exp. I, II, and III, and for Exp. IV, V, and VI, 54 *Ss* are available for determining the influence of distribution on learning the lists of low and high similarity. For the retention measurements, however, only 18 *Ss* are available for a given condition of distribution, similarity, and retention interval.

All *Ss* were given one practice day during which they learned a single list of paired-associates, partly by massing (4 sec. between trials), and partly by distribution (30 sec. between trials). This list was recalled after the same period used for that given experiment on the experimental days. During learning with 4 sec. between trials a single blank space appeared on the drum after each trial. For distribution at 30 sec. and 2 min., *S* named colors from a color board at the rate of one per sec. as timed by a metronome. In Exp. I and IV, in which the retention interval was 10 min., *Ss* were kept in the laboratory and practiced operating a motor division-of-attention test.

At recall all *Ss* were provided explicit instructions to give as many responses as possible on the first trial and to continue with the list until they could say it perfectly. Although *S* was allowed sufficient time to set himself for the recall, no formal warm-up task was given (7). Relearning in all conditions was by massed practice. In no case did *S* learn a new list on a day that he relearned a previous one.

Results

Learning as a function of distribution and intra-task similarity.—The influence of these two variables on rate of learning the lists to a criterion of eight correct responses is given in Table I. The results for each experiment are presented separately to demonstrate the consistency of the results from experiment to experiment.

The two major facts shown by the data in Table I are: (*a*) The high-similarity lists took much longer to learn than did the low-similarity lists; (*b*) No significant difference is present as a function of degree of distribution for the low-similarity lists nor for the high-similarity lists.

It is conceivable that while the number of trials required to reach a criterion of eight correct responses might not differ as a function of the degree of distribution, the rate of learning might change as a function of stage of learning. The learning curves for the low- and high-similarity lists for the three degrees of distribution have been plotted, and it is evident that little difference in the shapes of the curves exists.

Errors during learning.—Some evidence was presented in the previous paper in this series to the effect that if distributed practice facilitates learning, more errors will be present with distribution than with massing (11). In the present experiments it has been found that distribution does not facili-

TABLE I

THE INFLUENCE OF DISTRIBUTED PRACTICE AND INTRA-TASK SIMILARITY ON THE LEARNING OF PAIRED-ASSOCIATE LISTS

Type of List	Time Between Trials					
	4 sec.		30 sec.		2 min.	
	M	σ_M	*M*	σ_M	*M*	σ_M
Low Similarity						
Exp. I	10.50	.98	9.56	1.34	9.89	1.01
Exp. II	9.00	.74	8.83	.70	9.44	.68
Exp. III	10.11	1.18	9.50	1.26	10.28	1.34
All *Ss*	9.87	.56	9.30	.64	9.87	.56
High Similarity						
Exp. IV	16.61	1.29	16.78	1.37	15.67	1.04
Exp. V.	16.11	1.55	15.22	1.10	16.33	1.04
Exp. VI	14.44	1.39	14.33	1.19	17.50	2.30
All *Ss*	15.72	.81	15.44	.71	16.50	.90

tate learning. It may therefore be expected that the conditions—for a given level of similarity—will not demonstrate error differences. The data are presented in Table II in terms of mean number of errors per trial. To obtain these values the number of errors made by each *S* for a given condition was divided by the number of trials required to learn. The mean of these values for 54 *Ss* are the values given in Table II.

Table II shows that the frequency of errors was greater for the high-similarity lists than for the low, but that for a given degree of similarity the error frequency does not vary significantly for different degrees of distribution. The largest difference, that between 4-sec. distribution and 30- or 60-sec. distribution for high similarity, does not even approach statistical significance. While the error curves as a function of stage of learning will not be presented it can be reported that the shapes of these curves did not vary as a function of degree of distribution.

TABLE II

MEAN ERRORS PER TRIAL DURING LEARNING

Time Between Trials	Degree of Similarity			
	Low		High	
	M	σ_M	*M*	σ_M
4 sec.	1.00	.10	1.44	.09
30 sec.	.88	.10	1.61	.11
2 min.	.95	.11	1.61	.13

Recall.—Retention as measured by recall will be presented in terms of mean number of items forgotten. The values were determined for each *S* by subtracting the number of items recalled from the number of items given correctly on the last original learning trial. The means of these values for each condition are presented in Fig. 1 and Table III.

Looking first at the data for the low-similarity lists it is to be noted that, in general, forgetting is less following learning by distributed practice than following massed (4 sec.). At the 24-hr. retention point both of the distributed conditions show significantly less loss than the 4-sec. condition. The difference between the 4- and 30-sec. conditions is $1.22 \pm .51$

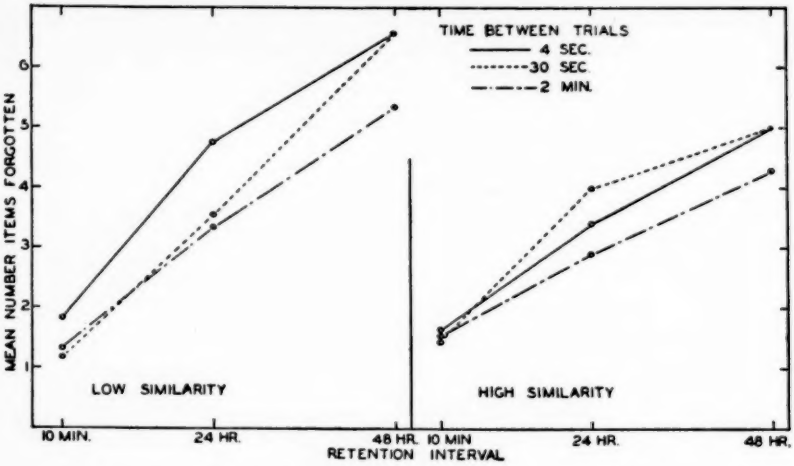

FIG. 1. Forgetting of paired adjectives as a function of intra-list similarity, length of interval between trials, and length of retention interval

items ($t = 2.39$). The difference between the 4-sec. and 2-min. conditions is $1.45 \pm .39$ ($t = 3.72$). For 17 *df* a t of 2.11 is significant at the .05 level of confidence and a t of 2.90 is necessary for the .01 level. After 48 hr. only the 2-min. condition of distribution yields better retention than the massed condition.

TABLE III

MEAN NUMBER OF ITEMS FORGOTTEN AS A FUNCTION OF TIME, INTRA-LIST SIMILARITY, AND DISTRIBUTION OF PRACTICE

Interval	Time Between Trials					
	4 sec.		30 sec.		2 min.	
	M	σ_M	*M*	σ_M	*M*	σ_M
Low Similarity						
10 min.	1.83	.34	1.22	.26	1.28	.29
24 hr.	4.78	.45	3.56	.34	3.33	.37
48 hr.	6.61	.38	6.61	.32	5.33	.44
High Similarity						
10 min.	1.67	.36	1.44	.28	1.61	.33
24 hr.	3.44	.42	4.00	.42	2.94	.42
48 hr.	5.00	.48	5.00	.45	4.28	.48

In contrast to the low-similarity lists, the high-similarity lists show little if any facilitation in retention following distribution. None of the differences between the massed and distributed conditions is significant. With the 2-min. distribution retention is consistently better than the massed condition, but clearly, as compared with the low-similarity lists, distribution of practice facilitates retention but little.

A third set of facts revealed is that high-similarity lists are retained somewhat better than low-similarity lists. According to Gibson's theory (**2**), forgetting should be greater the greater the amount of generalization within the list. Table III and Fig. 2 show this not to be the case. In Gibson's own experiment (**3**) it was found that high-similarity and low-similarity lists were retained about equally well after 24 hr. Her paired-associate lists had geometric forms for stimuli and

nonsense syllables as responses. In defense of her theory Gibson argues that the reason retention was unaffected by intra-list stimulus similarity was because the items of high-similarity lists had more repetitions during learning than did those of low-similarity lists (3, p. 189). In short, this implies that with equal number of reinforcements during original learning, retention will be better (forgetting less) for low-similarity lists than for high-similarity lists. This expectation will be tested next.

Item analysis of recall scores.—To make this test of the Gibson hypothesis the number of correct responses for each item in the original learning was tallied. For each item, "what happened" on the recall trial was also noted. If Gibson is correct, items appearing in a high-similarity list and having been given, say, three times correctly during original learning, should show greater forgetting than an item given correctly three times but appearing in a low-similarity list. This analysis was made for each condition of each experiment, but the consistency from condition to condition permits the data to be presented for all conditions of low similarity as opposed to all conditions of high similarity. Table IV gives the essential information. The per cent correct at recall and the per cent error at recall are plotted in Fig. 2.

TABLE IV

First-Trial Recall as a Function of Similarity when Number of Correct Responses during Original Learning is Held Constant

No. Correct during OL	Low Similarity		High Similarity	
	No. Cases	% Correct	No. Cases	% Correct
0	170	8.2	132	9.1
1	252	25.0	173	16.2
2	263	35.0	156	31.4
3	209	41.1	153	43.8
4	171	54.4	127	39.4
5	139	54.7	114	56.1
6	107	70.1	108	52.8
7	85	61.2	120	72.5
8	66	68.2	88	65.9
9	56	76.8	90	58.9
10	27	81.5	78	74.3
11 and over	75	80.0	281	78.3
Total	1620	44.5	1620	49.6

TABLE V

Total Number of Errors on First Recall Trial

Interval	Time between Learning Trials			Total
	4 sec.	30 sec.	2 min.	
Low Similarity				
10 min.	15	10	21	46
24 hr.	15	16	17	48
48 hr.	11	12	19	42
High Similarity				
10 min.	30	28	26	84
24 hr.	17	22	23	62
48 hr.	27	28	38	93

In terms of per cent correct at recall following a constant number of reinforcements (correct responses) during original learning, there is little evidence to support the Gibson theory. *When equated for number of correct responses during original learning low- and high-similarity lists are recalled about equally well.* The fact that Table III showed retention to be somewhat better for the high- than for the low-similarity lists is probably to be accounted for by the high reinforcement (7, 8, 9, 10, 11 and over) given the high-similarity lists as compared with the low-similarity lists.

With respect to the percentage error for each level of reinforcement, two facts are clear: (a) as degree of learning increases, number of errors decreases, and (b) more errors are made in recalling the high-similarity

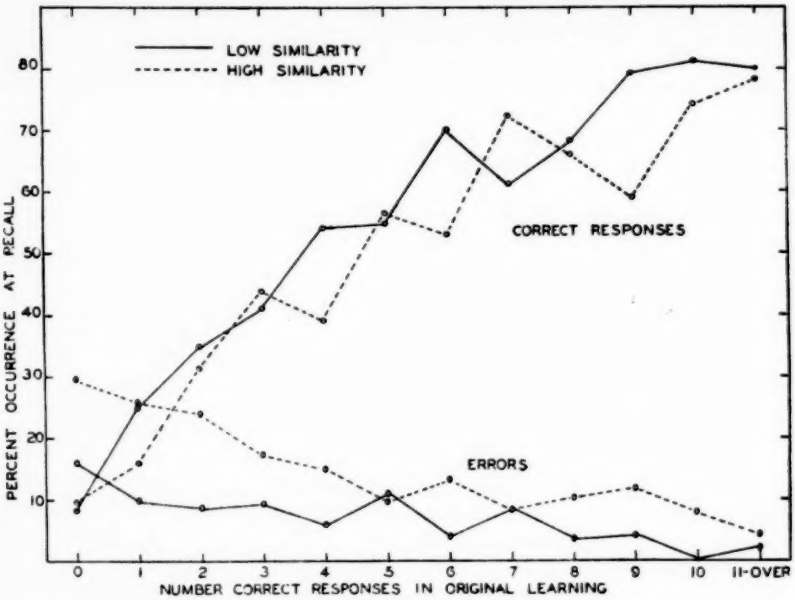

FIG. 2. Correct responses and errors at recall for low- and high-similarity lists when number of correct responses during original learning is held constant (all retention intervals combined)

lists than in recalling the low-similarity lists. Thus, while there are more errors made in recalling the high-similarity lists—as would be predicted by the Gibson theory—these errors apparently do not reduce the effectiveness of the recall of correct responses. The examination of errors leads to another test of Gibson's theory. If generalization recovers as a function of time, its recovery should be indexed by an increase in the number of errors as the retention interval increases. Data for evaluating this deduction are given in Table V.

The data of Table V do not confirm the idea that errors increase as a function of time since learning. On the basis of these data it is more correct to say that errors remain roughly constant as a function of passage of time. The variation in total errors exhibits no regular trend.

Relearning.—The *S*s relearned the list under each condition to a criterion of one perfect trial. These data will not be presented. In general, more trials to relearn were required as the retention interval increased. Relearning was always faster following distributed practice than following massed practice, but in no case were the differences significant. Errors per trial were greater for the high- than for the low-similarity lists.

DISCUSSION

The intervals of distributed practice used here failed to facilitate the learning of paired-associate lists of either high or low similarity. And, although the error frequency was

greater for lists of similar items than for dissimilar items, no difference in frequency of errors was present as a function of degree of distribution of practice. Thus, as in the first study in this series (11), facilitation by distributed practice was not found when the error frequency was the same for massed and distributed practice.

Considering past studies as a whole, it may be said that distributed practice has been found to benefit learning when difficulty is high. Difficulty may be increased by stepping up the presentation rate, by lengthening the list, or by increasing intra-task similarity. It was expected that the high-similarity lists used in the present experiments would result in increased rate of learning under distribution. But, even though the task was difficult—several *Ss* had to be dropped because they could not learn the lists—no difference in learning under massing and distribution was observed. Increased rate of presentation of the high-similarity lists might accomplish the objective. Hovland (5) has recently shown that if speed of presentation of paired nonsense syllables was increased from a 2-sec. rate to a 1-sec. rate, distribution of trials benefited learning. Nevertheless, the fact remains that under the conditions of the present experiments, no effects of distributed practice in learning were observed.

When the experiments were planned it was believed that distributed practice would allow for successive extinction of error tendencies. A rest interval, it was presumed, would allow recovery of intra-list generalization tendencies. On the next trial these tendencies would be differentiated or extinguished again. With the next rest interval the tendencies might recover but would recover less than previously; hence, they would be extinguished more rapidly on the next trial, etc. The net effect would be very slow forgetting following distributed practice because the error tendencies, having had repeated extinctions, would recover very slowly. There are at least three sets of facts which make this hypothesis untenable. (a) Retention of high-similarity lists should benefit more by distributed practice than the low-similarity lists because of the greater number of generalization tendencies. The reverse of this was found. (b) Errors during learning should be greater for distributed than for massed practice, at least during the early stages of learning. The data on errors as a function of stage of learning were not presented, but it can be categorically stated that they did not vary as a function of interval between trials. (c) Finally, if the hypothesis is tenable, there should be fewer errors at recall following distribution than following massing. This was not the case; indeed, there tended to be a few more errors following distribution (see Table V). In short, the idea of successive extinctions to account for superiority of retention following distributed practice must be rejected on the basis of the data. Other explanations will be considered.

A theory which postulates a form of reactive inhibition which accrues during massed learning and not during distribution cannot account for the better retention following distribution. Such a theory would predict better retention following massed practice than following distribution. The Yale theory of rote learning (6) reached an impasse on this matter. Likewise, McClelland's performance-inhibition theory (9), since it is similar to the Yale theory on this point, would not be able to account for the retention data of the present experiments.

Another possibility might be stated in terms of practice at recalling. Cain and Willey (1) offer this as one possible explanation for better retention following distribution. The *S*, having learned by distribution, has considerable practice in recalling during learning. He would not have such practice during massing. To account for the present data such a theory would have to be stated in terms of specific lists; that is, it would have to state that *S* had specific practice in recalling a given list since all *S*s had a great deal of practice in recalling following rest intervals. Yet, even with this added assumption, there still remains the problem of explaining why distribution was not effective in the recall of the high-similarity lists. The idea of "practice in recalling" does not seem to be a fruitful approach.

At the present time no satisfactory theory to account for the retention results as a function of distribution can be advanced.

Our results have been interpreted as being contrary to the expectations of Gibson's theory (2). Three findings may be so interpreted: (a) equally good retention of items in high- and low-similarity lists when equated for degree of original learning; (b) failure of errors to increase as a function of retention interval, and (c) the failure of the greater number of errors at recall for high-similarity lists (as compared with low-similarity lists) to be correlated with greater forgetting. In fairness to Gibson it should be pointed out that her theory is stated only in terms of stimulus generalization. The high-similarity lists used here were constructed so as to produce stimulus generalization, response generalization, and stimulus-response generalization. It is conceivable that the construction of these lists was such that other forms of generalization resulted in processes which masked or nullified the operation of stimulus generalization as postulated in the Gibson theory. This does not seem too likely, however, in view of the recent findings that response generalization seems to have the same properties as stimulus generalization (8, 10, 12). Nevertheless, before rejecting the Gibson postulates concerning retention they should be tested under the conditions she had in mind when formulating the theory. We have found no confirming evidence and some which may be only apparently contrary to the theory.

We have said that errors did not increase as a function of the retention interval as predicted by Gibson's theory. Although Gibson has not suggested it, if we took the ratio of correct responses to errors it would be shown that relative to correct responses, overt errors increase as a function of the retention interval. Gibson's theory does not, of course, say that errors should increase in frequency indefinitely. At some point they will start falling in frequency. No evidence of this fall is present with the intervals used here.

With regard to the observation that retention of low- and high-similarity lists was equally good, it might be argued that a longer retention interval would eventually have produced greater forgetting for the high-similarity lists. However, the only trend noticeable was in the opposite direction; the low-similarity lists tended to be retained relatively better after the short interval than after the 48-hr. period.

Summary

These experiments were designed to determine the influence of (a) distributed practice, and (b) intra-list similarity on learning and retention of lists of ten paired adjectives. The

three degrees of distribution were 4 sec., 30 sec., and 2 min. between each learning trial. There were two degrees of intra-list similarity, one very high and one low. The three retention intervals were 10 min., 24 hr., and 48 hr. The 18 conditions were run as six experiments of three conditions each, with 18 *Ss* in each counterbalanced experiment. The basic results follow:

1. There were no differences in rate of learning to the criterion of 8 out of 10 correct responses as a function of length of interval between trials.
2. The high-similarity lists took many more trials to learn than did the low-similarity lists. More errors per trial were also evident.
3. Distributed practice facilitated retention for the low-similarity lists, but not for the high-similarity lists.
4. The frequency of errors at recall showed no relation to the length of the retention interval.
5. When equated for degree of original learning, items in the low- and high-similarity lists were retained equally well.
6. The applicability of several theories of verbal learning and retention to these results was evaluated. These theories included the reactive inhibition-type theory of Hull and his coworkers; McClelland's performance-inhibition theory, and Gibson's generalization hypothesis. None of these theories was found capable of handling the basic findings of the present study.

(Manuscript received
September 14, 1950)

References

1. Cain, L. F., and Willey, R. de V. The effect of spaced learning on the curve of retention. *J. exp. Psychol.*, 1939, **25**, 209–214.
2. Gibson, Eleanor J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
3. Gibson, Eleanor J. Intra-list generalization as a factor in verbal learning. *J. exp. Psychol.*, 1942, **30**, 185–200.
4. Haagen, C. H. Synonymity, vividness, familiarity, and association-value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27**, 453–464.
5. Hovland, C. I. Experimental studies in rote-learning theory: VIII. Distributed practice of paired associates with varying rates of presentation. *J. exp. Psychol.*, 1949, **39**, 714–718.
6. Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., and Fitch, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale Univ. Press, 1940.
7. Irion, A. L. The relation of "set" to retention. *Psychol. Rev.*, 1948, **55**, 336–341.
8. Morgan, R. L., and Underwood, B. J. Proactive inhibition as a function of response similarity. *J. exp. Psychol.*, 1950, **40**, 592–603.
9. McClelland, D. C. Studies in serial verbal discrimination learning: II. Retention of responses to right and wrong words in a transfer situation. *J. exp. Psychol.*, 1942, **31**, 149–162.
10. Underwood, B. J. Response similarity and degree of first-list learning in associative facilitation and inhibition. *Amer. Psychologist*, 1950, **5**, 275. (Abstract)
11. Underwood, B. J., and Goad, D. Studies of distributed practice: I. The influence of intra-list similarity in verbal learning. *J. exp. Psychol.*, 1951, **42**, 125–134.
12. Underwood, B. J., and Hughes, R. H. Gradients of generalized verbal responses. *Amer. J. Psychol.*, 1950, **63**, 422–430.

THE EFFECT OF MOTIVATION AND AMOUNT OF PRE-REST PRACTICE UPON INHIBITORY POTENTIAL IN MOTOR LEARNING[1]

BY HILTON N. WASSERMAN

George Fry and Associates[2]

The immediate background of the present experiment may be best understood by referring to Kimble's (7) summary and extension of Hull's performance inhibition theory (3). According to this theory, the total inhibitory potential is made up of two elementary decremental components, reactive inhibition (I_R) and conditioned inhibition ($_sI_R$). These two factors are alike in that they both depress the habit being learned. They differ, however, in many ways.

Reactive inhibition is a drive state closely allied to pain avoidance. It is response-produced inhibition which results from or accompanies all effortful behavior - whether reinforced or not, and dissipates during a rest interval. If the rest intervals are not long enough to allow for the dissipation of a good deal of the I_R (as in massed practice), I_R accumulates and the need for rest increases. When resting occurs, I_R dissipates and this serves as a drive reduction or reinforcement.

Because of the reinforcement caused by rest, the response of resting will become a habit which will be conditioned to whatever stimuli are present in the learning situation. This conditioned resting tendency is $_sI_R$.

The two inhibitory components are, then, a drive (I_R) and a habit ($_sI_R$), and each must have the characteristics, respectively, of other drives and habits. Thus, $_sI_R$ should develop as a positive growth function of the number of reinforcements, while I_R might be expected to accumulate as some increasing function of the amount of effort expended. Further, conditioned inhibition, being a habit, should show almost no diminution with the passage of time, while I_R, being a drive, will be reducible by the appropriate goal response, resting.[3] This latter difference makes it possible (as has been done by Kimble, 7, p. 18) to trace the separate course of development of each of these two inhibitory components.

Kimble (7) hypothesizes that with continuous practice (work), since I_R is a drive, the accumulation of a certain critical amount will automatically produce resting, thus resulting in dissipation of I_R before a formal rest interval is introduced. If motivation to perform the task continues, once I_R is reduced below the critical level, the organism will resume work and continue until the critical level of I_R is reached again. Hence, a sort of

[1] This paper constitutes a portion of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, in the Department of Psychology in the New York University Graduate School of Arts and Science. The writer wishes to acknowledge with sincere thanks the unfailing aid and advice provided by Professor Leland W. Crafts throughout the progress of this investigation. The writer is also grateful to Professor Howard H. Kendler for his valuable suggestions, and to Professor Thomas N. Jenkins for his suggestions concerning the statistical treatment of the data.

[2] This study was completed while the author was at New York University.

[3] It should be pointed out, however, that because $_sI_R$ is a habit, it should be affected by retroactive inhibition. Hence, what happens during an interval of time affects $_sI_R$, while the mere passage of time (resting) after the cessation of the response diminishes I_R.

equilibrium will be reached in which the organism rests long enough to keep I_R at, or slightly below, some specific level. It is to be expected that whenever motivation decreases in learning, this level will drop somewhat.

Kimble further hypothesizes that some threshold value of I_R is necessary before the process of development of ${}_sI_R$ can begin. Thus, early in learning, when I_R has not yet reached this threshold value, no ${}_sI_R$ should be present.

The results of Kimble's experiment (7) support these two hypotheses and suggest a third: Even though the curve of development of ${}_sI_R$ tends toward negative acceleration (as is expected for a habit), Kimble suggests that the reaching of the asymptotic value will be postponed because any condition which tends to reduce the tolerable (critical) amount of I_R will increase the amount of resting, and therefore, the rate of development of ${}_sI_R$. One such condition tending to reduce the critical amount of reactive inhibition could be the decrease in motivation which occurs late in learning with certain activities.

By utilizing two motivation conditions, each having a spaced practice control group, a massed practice control group, and nine experimental groups working under conditions of massed practice except for the introduction of a 10-min. rest after varying amounts of practice, the present experiment is designed to test the following hypothesis and five predictions related to it. The hypothesis is that high motivation should raise the critical (tolerable) level of I_R necessary to produce the automatic resting response. The predictions based upon this hypothesis and the preceding discussion are as follows:

1. Since high motivation should raise the organism's resistance to I_R, i.e., permit the accumulation of a higher level of I_R before an automatic resting response occurs, and since this resistance might, in turn, involve a greater expenditure of effort, thus resulting in more I_R, it may be predicted that late in learning I_R should be less with a poorly motivated group than with a highly motivated group. Hence, a rest interval late in learning should be more beneficial (in terms of dissipation of I_R) for a highly motivated group than for a poorly motivated group.

2. If the amount of improvement after rest (which is due to the dissipation of I_R during rest) becomes smaller the later in learning the rest interval is introduced, as Kimble (7) and others have found, this decrease in improvement after rest should be less for the highly motivated group than for the poorly motivated group.

3. Early in learning, before the threshold value of I_R has been reached, no ${}_sI_R$ should be present in either of the motivation conditions.

4. Since with the highly motivated group there will be less resting late in learning, and therefore, less frequent reinforcement (need reduction), it might be expected that late in learning the rate of development of ${}_sI_R$ should be less in the highly motivated group than in the poorly motivated group, i.e., the asymptote should be reached sooner with high motivation than with low motivation.

5. The amount of ${}_sI_R$ late in learning should be less in the highly motivated group than in the poorly motivated group. This prediction follows from the fact that even though less I_R will accumulate in the poorly motivated group, the need reduction (reinforcement of ${}_sI_R$) produced by resting will be more frequent in this group than in the highly motivated group.

Method

A total of 22 groups of *Ss* was utilized, each group being a class in general psychology at Washington Square College of New York University. The experiment was administered to one group at a time, and a total of 538 *Ss* was used.

The learning task was the alphabet printing task described by Ruch and Warren (**12**), and utilized by Kimble (**7, 8**) and Kientzle (**6**). Briefly, the task requires *S* to print the letters of the alphabet in normal A to Z order, but upside down and from right to left so that when the work sheet is rotated 180 degrees, the letters appear properly oriented.

With the exception of the motivation variable, the procedure was essentially that used by Kimble (**7**). The *Ss* were divided into two motivation conditions, low and high. Each motivation condition was composed of 11 groups of *Ss*: a spaced control, a massed control, and nine experimental groups. Each of the experimental and control groups had 20–30 *Ss*. The nine experimental groups were given 50 massed 30-sec. trials and a single 10-min. rest after trials 2, 5, 10, 15, 20, 25, 30, 35, and 40, respectively. The control groups were given 50 30-sec. trials, the massed groups receiving no rest and the spaced groups receiving a 30-sec. rest after each trial.

Each *S* was provided with five mimeographed work sheets and a pencil. The work sheets contained space for ten trials. To assure that all *Ss* printed letters of about the same height and width, the work sheet was set up in the form of cross section paper 4 × 4 to the inch so that each letter had to be printed in a ¼-in. square box. The five sheets were stapled together so that the *Ss* were required to turn a page after trials 10, 20, 30, and 40.

The task was explained and demonstrated to the *Ss* in both motivation conditions and they started working at the signal "go." In each group except the spaced control groups the *Ss* were told after 29 sec. to go on to the next trial.[4] Since the *Ss* were required to turn a page after trials 10, 20, 30, and 40, a little time was lost on trials 11, 21, 31, and 41. Hence, the scores on these four trials were discarded except in those groups which had rest after trials 10, 20, 30, and 40. In the massed control groups the *Ss* were given a simple "stop" and "go" set of signals.

Following Kientzle's procedure (**6**), during the rest interval the *Ss* were told to remain "at ease" in their seats, and were asked not to discuss the work with their neighbors. Also, following Kientzle, the score was the number of letters printed on each trial regardless of accuracy.

The difference in motivation was provided by the instructions given in the two conditions. The low motivation condition was task-oriented, the *Ss* being told that they were merely helping out with some preliminary experimentation to determine if the task could be adapted to a proposed future experiment. The high motivation condition was ego-oriented, the *Ss* being told that they were going to be given a new type of intelligence test the results of which would be reported to them and held strictly confidential. These *Ss* were further told that the test was being given to the various psychology classes in order to set up average levels of performance.

At the end of the experimental session the *Ss* in both motivation conditions were asked to indicate on the back of their work sheets any pencil difficulties or anything else that might have interfered with their performance.

Results and Discussion

Learning.—The basic data of the experiment are shown in Fig. 1 and 2. The upper and lower curves show the performance of the spaced control groups and the massed control groups respectively. The intermediate curves show the post-rest performance of the nine experimental groups (in each motivation condition) on the first ten trials following a 10-min. rest. The number on each of these curves is the number of the trial after which rest was given. The massed practice curves include the performance of all the experimental groups before the introduction of the rest. The *N*s for the various parts of the low-motivation massed practice curve (Fig. 1) are as follows: Trials 1–2, 242; Trials 3–5, 222; Trials 6–10, 199; Trials 12–15, 169; Trials 16–20, 141; Trials 22–25, 119; Trials 26–30, 94; Trials 32–35, 66; Trials 36–40, 46; Trials 42–50, 26. The comparable *N*s for the various parts of the high-motivation massed practice curve (Fig. 2) are as follows: 245, 221, 196, 168, 143, 116, 96, 75, 51, 30.

It should be further noted that in plotting the massed practice curves

[4] See Kimble (**7**, p. 17) for the rationale of this timing of the trials.

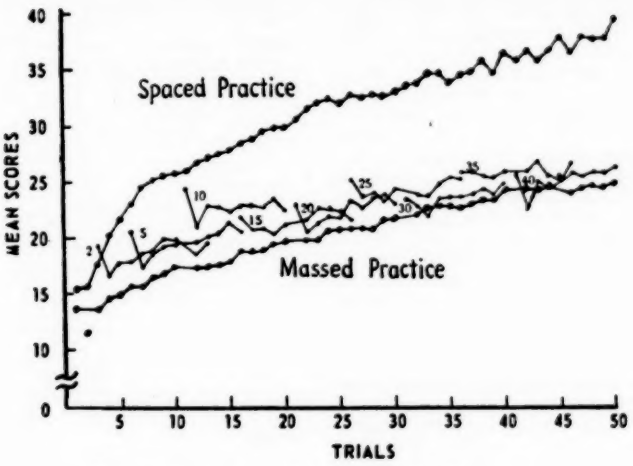

FIG. 1. Learning curves for the low motivation groups

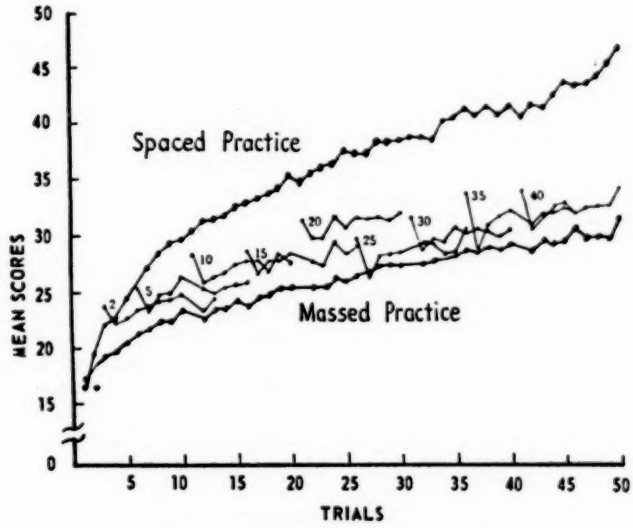

FIG. 2. Learning curves for the high motivation groups

shown in Fig. 1 and 2, Trials 11, 21, 31, and 41 were omitted because they were shortened by the necessity of turning pages of the work booklets. In addition, the point for Trial 2 is shown on the graphs, but was not utilized in drawing the curves because it was apparent that most of the *Ss* became confused in making the transition from Trial 1 to 2 and as a result did poorly on Trial 2 (as shown by the point on the graphs).

It may be determined from Fig. 1 and 2 whether the results bear out two basic assumptions which the design of the experiment necessitated. First, following Kimble (**7**), it was assumed that the massed practice control groups worked under conditions of nearly maximal inhibition, while inhibition was near a minimum for the spaced practice control groups. It is quite evident from Fig. 1 and 2 that the spaced groups are much superior to the massed groups, the difference being greatest on the last trial of the experiment in both motivation conditions. This finding would be expected if the accumulation of inhibition with massing produced the difference. Comparison of the present findings with those of Kimble (**7**) further support this assumption. The results of these two experiments are quite parallel in terms of performance scores with Kimble's groups falling between the low and high motivation conditions of the present experiment. Thus, in the light of these comparisons and the finding by Kientzle (**6**) that diminishing returns (in terms of amount of reminiscence) set in after 45 sec. of rest with *1-min.* trials, this first basic assumption seems quite justified.

A second basic assumption inherent in the design of this experiment is that the performance of the high motivation groups would be superior to that of the low motivation groups. Preliminary experimentation with several sets of instructions revealed that it was not a matter of getting a maximum amount of motivation, but rather an optimum amount. It is quite evident from Fig. 1 and 2 that with the instructions which were finally utilized the performance of the high motivation groups was consistently superior to that of the low motivation groups. Further, these differences were significant at least at the .05 level of confidence, and in many cases beyond the .01 level of confidence.

Reminiscence (I_R).—It may be noted from Fig. 1 and 2 that a striking feature of each of the experimental groups under both motivation conditions is a marked increase in performance after rest (reminiscence). Table I presents the mean increase for each of the 18 experimental groups along with the respective *t*-values of these increases.

The occurrence of reminiscence provides a method of measuring reactive inhibition. It was here assumed (as other investigators have) that reminiscence is produced by the dissipation of I_R during rest. Thus, if the rest interval is long enough to allow for the dissipation of all or most of the accumulated I_R, the magnitude of the increase after rest provides an indirect measure of the amount of I_R that was present before rest.

TABLE I

MEAN INCREASE IN SCORE AFTER REST FOR EACH EXPERIMENTAL GROUP

Motivation	Last Trial before Rest	*N*	Mean Increase	*t*
Low	2	20	7.10	12.46**
	5	23	5.35	7.23**
	10	30	7.60	10.70**
	15	28	4.18	5.36**
	20	22	4.63	8.12**
	25	25	4.56	9.31**
	30	28	2.89	5.45**
	35	20	1.95	3.05*
	40	20	2.05	1.92
High	2	24	7.29	9.85**
	5	25	5.16	7.48**
	10	28	5.39	9.14**
	15	25	4.96	9.19**
	20	27	5.48	7.83**
	25	20	4.65	7.88**
	30	21	5.15	9.36**
	35	24	4.96	8.27**
	40	21	5.14	6.85**

* Significant at the .01 level of confidence.
** Significant at the .001 level of confidence.

The assumption that the 10-min. rest utilized in this experiment was ample to allow for nearly complete recovery from inhibition is supported by Kimble's studies with the alphabet printing task (**7, 8**) and also by the earlier pursuit rotor studies of Ammons (**2**), Kimble and Horenstein (**10**), and Irion (**5**), all of which suggest that the maximal rate of dissipation of reactive inhibition takes place during the first few minutes of rest with greatly diminished returns after about 5 min. of rest.

One additional preliminary point remains to be dealt with before proceeding with the evaluation of the reminiscence data. This point concerns the statistical method used in dealing with

reminiscence in this experiment. Usually the reliability of the increase after rest is determined by comparing the increase in the experimental group with the " normal" trial-to-trial increase which occurs in the massed control group without rest. In the present experiment, however, this comparison could be made only with those groups for which comparable massed practice scores were available. These scores were not available for Trials 11, 21, 31, and 41 because of the necessity of turning pages on these trials. Thus, the "normal" increase could not be computed for Groups 10, 20, 30, and 40.

The substitute procedure which was used in this experiment (and apparently by Kimble [7]) assumed the pre-rest trials of the experimental groups to be comparable to the corresponding trials of the massed practice group. The null hypothesis was then set up that the introduction of a rest interval will produce no special increase in score, i.e., no increase other than the "normal" trial-to-trial increase. The p-values in Table I, of course, lead to the rejection of the null hypothesis.

That this procedure is valid is confirmed by the p-values presented in Table II which are based upon the five comparisons with the massed control group which could be made since the turning of pages was not necessary at these points. Similar trends are indicated by both methods, there being only three points where the trends shown by Tables I and II are not in agreement. Furthermore, similar results were also gotten by use of the chi-square technique. Chi-square analyses confirmed the reliability of the increases after rest under both conditions of motivation, and also the reliability of the differences between the two motivation conditions (which will be discussed below).

TABLE II

MEAN DIFFERENCES BETWEEN INCREASES IN SCORE AFTER REST AND INCREASES WITHOUT REST *

Motivation	Last Trial before Rest	Mean Diff.	t
Low	2	3.33	3.33**
	5	4.23	4.03***
	15	3.37	3.30**
	25	5.06	6.25***
	35	1.37	1.69
High	2	4.89	5.04***
	5	4.29	4.42***
	15	5.20	7.22***
	25	4.02	5.74***
	35	3.56	3.71***

* Groups having rest after trials 10, 20, 30, and 40 not usable (see text).
** Significant at the .01 level of confidence.
*** Significant at the .001 level of confidence.

The results of the present experiment confirm and supplement Kimble's (7) findings concerning the development of reactive inhibition. First, a major gap in Kimble's experiment may be filled in by the present findings. The use of a group having rest after Trial 2 was designed specifically to test the hypothesis that I_R builds up extremely rapidly. The data presented in Table I lend support to this hypothesis. Also, Kimble found that the maximum amount of I_R was present after 7.5 min. of work, with 88.5 per cent of this maximum being present after 2.5 min. of work, and 99.4 per cent of it being present after 5 min. of work. The results of the present experiment show that with low motivation the maximum amount of I_R was reached after 5 min. of work, with 93.4 per cent of this maximum being present after only 1 min. of work. With high motivation the maximum was reached after only 1 min. of work, and the level of I_R was maintained throughout learning at about 70 per cent of this maximum.

The percentages of the present experiment are probably somewhat inflated by a complicating factor. It was noted above that there was a decrease in score on Trial 2 due to confusion on the part of the *Ss* in making the transition from Trial 1 to 2. Since the difference between the last pre-rest trial and the first post-rest trial was the measure of reminiscence used, the actual amount of reminiscence in the two groups having rest after Trial 2 was probably less than the value shown in Table I. That is, part of the difference was due to an increase after rest produced by the dissipation of I_R, but also part of it must have been due to the prior decrease in score on Trial 2. This inflation of the measure may be dealt with by using the score on Trial 1 rather than the (lower) score on Trial 2 to calculate the increase after rest in these two groups. When this is done, the amount of reminiscence for Group 2 with low motivation becomes 6.15 instead of 7.10, and with high motivation 6.59 instead of

7.29. It is obvious that the indicated trend is but little changed by this correction (actually overcorrection). With low motivation 80.9 per cent of the maximum amount of I_R was present after only 1 min. of work, while with high motivation the maximum was still reached after only 1 min. of work, with the level of I_R being maintained throughout learning at about 77.5 per cent of this maximum.

Regardless of the *exact* point at which the maximum I_R was reached, it seems justified to conclude that in the present study I_R was built up extremely rapidly. These findings are in substantial agreement with those of Ammons (**2**) who found with pursuit learning that 3 min. of work produced almost as much reminiscence as the maximum amount which was reached after 8 min. of work.

On the basis of our main hypothesis two major predictions concerning I_R were made. The first prediction was that late in learning there should be less I_R with low motivation than with high motivation. This prediction is confirmed by the data presented in Table III. It may be seen that the differences between the two motivation conditions are not of any great magnitude and are not statistically significant (with one exception) until late in learning (after 30 pre-rest trials). Hence, the inference that high motivation should result in less resting *during* learning and thus, more I_R late in learning is clearly substantiated by the results of this experiment. Apparently, the poorly motivated *Ss* more frequently went into an automatic resting response resulting in the dissipation of I_R *during* learning, while the highly motivated *Ss* resisted this tendency to a greater extent.

These findings are in agreement with those of Kimble (**9**) who used the pursuit rotor and a man-to-man competitive situation to obtain two levels of motivation. He hypothesized that *Ss* who are unsuccessful in the competitive situation are more highly motivated to do well in the learning task than successful *Ss*. He then made the prediction that the unsuccessful *Ss* will tolerate more reactive inhibition (will show more reminiscence after a rest interval) than the successful *Ss* because the former are more highly motivated. His results confirmed this prediction, the overall difference in reminiscance being statistically significant at the .02 level of confidence.

TABLE III

AMOUNT BY WHICH THE MEAN INCREASE IN SCORE AFTER REST WITH HIGH MOTIVATION EXCEEDS THAT WITH LOW MOTIVATION

Last Trial before Rest	Mean Difference	t
2	.19	.20
5	−.19	.19
10	−2.21	2.21*
15	.78	.82
20	.85	.91
25	.09	.12
30	2.26	2.31*
35	3.01	3.42**
40	3.09	2.38*

* Significant at the .05 level of confidence.
** Significant at the .01 level of confidence.

Closely related to these findings is the second prediction concerning I_R. Since the amount of reminiscence tends to decrease as the rest interval is introduced progressively later in learning, it was predicted that this decrease should be less with high motivation than with low motivation. The results shown in Tables I and II confirm this prediction. It may be seen in Table I that with high motivation, after an initial drop (from Group 2 to 5), the level of I_R fluctuates up and down a little, maintaining a fairly constant level. However, with low motivation there is a definite tendency for the level of I_R to decrease from the maximum point as the amount of pre-rest practice increases.

The partial results shown in Table II also support the prediction, although not as clearly. Late in learning (from 25 to 35 pre-rest trials) with low motivation there is a decrease in I_R of 3.69, while with high motivation there is a decrease of only .46.

An additional point concerning the warming-up effect may be noted. As discussed by Ammons (1, 2) and Irion (5), the magnitude of the warming-up effect is indicated by the amount by which the second post-rest trial exceeds the first. Using this method of measurement, Kimble (7) notes that his data fail to exhibit the warm-up phenomenon which Ammons (2) demonstrated with the pursuit rotor. That is, Kimble's post-rest performance curves show a drop on the second post-rest trial, and, therefore, no warming-up effect is exhibited. This is especially interesting since the results of the present experiment, using the alphabet printing task as did Kimble, also show a rather sudden drop on the second post-rest trial. This is clearly evident in Fig. 1 and 2.

It seems possible that Kimble's data and those of the present experiment failed to exhibit any decrement due to need for warm-up because this decremental influence was greatly outweighed by the beneficial effect of the dissipation of I_R during rest. It is also possible that in these experiments the warming-up effect was not of considerable magnitude (especially with the smaller amounts of pre-rest practice) since in both only a 10-min. rest was used, while Ammons found the maximum warm-up decrement after 20 min. of rest in pursuit learning.

If this explanation is correct, the increase on the first trial after rest must have been diminished somewhat by the opposite influence of the need for warming-up. This inference has been tested by Irion (4), with nonsense syllables, by introducing varying amounts of warming-up activity just before the resumption of practice after a rest. He found that warming-up activity was beneficial up to a point, but then apparently produced additional reactive inhibition which diminished recall. With these results in mind, the introduction of varying amounts of warming-up activity just before the resumption of work should result in an increase in reminiscence up to a point, and then a decrease as the amount of warming-up activity increases.

Results relevant to this point are provided by one of Kientzle's groups (6) in which ten 1-min. trials on the alphabet printing task were separated by 1-min. rest periods. This group was unique in that the entire rest interval was spent in making tallies and cross bars on a separate sheet of paper. From the above discussion, it would be expected that with a 1-min. work period and a 1-min. rest the optimal amount of warming-up activity would be very small, and that 1 min. of warming-up activity would probably result in a decrement rather than facilitation. This group in Kientzle's study *was* inferior in learning to another group differing from it only in that the rest intervals were spent in resting.

A recent study by Silver (13) was designed specifically to investigate this problem. He utilized the alphabet printing task and introduced varying amounts (0, .5, 1, 1.5, 2, and 2.5 min.) of massed and distributed warming-up activity during a 10-min. rest following five, fifteen and thirty 30-sec. trials. The warming-up activity was that used by Kientzle. He found that *Ss* who received warming-up activity were superior on the first post-rest trial to those who had rest with no warming-up activity. In addition, with increasing amounts of massed warming-up activity, reminiscence increased up to a point and then declined. On the other hand, with increasing amounts of distributed warming-up activity, reminiscence progressively increased.

These findings are in agreement with the inferences stated above, and serve to clarify, at least partially, the relationship between reminiscence and warming-up effect.

Permanent decrement ($_sI_R$).—The previous discussion has noted some of the problems inherent in the use of reminiscence as a measure of I_R. Additional problems are encountered in the measurement of $_sI_R$.

Kimble (7) has advanced one method of measuring $_sI_R$ which makes use of the assumption that I_R, being a drive, dissipates during rest, while $_sI_R$, being a habit, does not dissipate with time. Thus, whatever decrement remains after a rest interval sufficient to dissipate I_R completely, should be due to $_sI_R$, and the measurement of this "permanent" decrement may be used as an indirect measure of $_sI_R$. Further, since in Kimble's experiment (and in the present one) the spaced practice group is assumed to be subject to a minimum of inhibition, the amount by which the performance of the spaced group exceeds that of the experimental groups on each initial post-rest trial should give an indication of the amount of permanent decrement, or $_sI_R$, that was present in the experimental groups. Kimble's procedure was to subtract the score on the first post-rest trial of each experimental group from the score of the spaced group on the trial of the same ordinal number.

However, part of the superiority of the spaced group may not be due to the decremental effects of $_sI_R$ on the experimental group, but rather to the beneficial effects of a greater number of reinforcements (greater habit strength) in the spaced group. This superiority of spaced practice is indicated by the higher level of performance attained on each trial, and becomes greater as practice progresses. Thus, the spaced group, at almost every point in the learning process, has printed a greater total number of letters than have the experimental groups (which have massed practice except for the single rest interval). It follows that the spaced group has received more frequent reinforcement, and therefore, habit strength should be greater in the spaced group than in the experimental groups at every trial except the first (3). Therefore, any difference between the initial post-rest performance of the experimental group and the spaced group on the trial of the same ordinal number confounds greater habit strength ($_sH_R$) in the spaced group and conditioned inhibition ($_sI_R$) in the experimental group.

In view of this criticism of Kimble's method of measuring $_sI_R$, we have used, in addition to this method, a modification termed the cumulative mean method, which attempts to take into account the factor of differential number of reinforcements by comparing the experimental and spaced groups on trials following similar numbers of responses (letters printed) rather than on trials of the same ordinal number. Although the matching of the cumulative means could only be approximate, the results are consistent and clear-cut.

It may be seen from Table IV that the two methods suggest very similar trends, the main difference between the results of the two being that the values of $_sI_R$ obtained by the cumulative mean method are consistently smaller than those obtained by Kimble's method. It seems probable, then, that Kimble's measure *is* being contaminated somewhat by some additional factor (perhaps habit strength). However, since only the absolute magnitude of the values is affected, while the relationships between the values remain approximately the same, the validity of Kimble's conclusions concerning $_sI_R$ is not affected.[5]

Three specific predictions concerning the development of $_sI_R$ have been made. According to the first, before the threshold value of I_R has been reached, no $_sI_R$ should be present, i.e., no $_sI_R$ should be present very early in learning. Further, since high motivation results in raising the critical level of I_R, and thus, less resting *during* learning, it is expected that the initial appearance of $_sI_R$ occurs somewhat later with high motivation than with low motivation. It may be seen from Fig. 1 and 2, and from Table IV, that the results of this experiment confirmed every aspect of this prediction.

In Kimble's experiment (7), $_sI_R$ first appeared after ten pre-rest trials. If the motivation in Kimble's experiment is assumed to be somewhat greater than that of the low motivation groups of the present experiment

TABLE IV

AMOUNT BY WHICH THE SPACED GROUP EXCEEDS THE REMINISCENCE POINT AFTER NINE AMOUNTS OF PRE-REST PRACTICE

Motivation	Group	Kimble's Method		Cumulative Mean Method	
		Mean Diff.	t	Mean Diff.	t
Low	2	−1.56	—	−1.56	—
	5	2.55	1.83	1.31	0.94
	10	1.84	1.64	1.21	1.01
	15	6.35	5.57***	4.66	3.88***
	20	7.39	5.77***	4.81	4.04***
	25	7.59	7.09***	4.21	5.77***
	30	10.06	10.06***	7.71	8.86***
	35	8.48	7.50***	6.79	5.96***
	40	10.24	8.26***	7.14	6.16***
High	2	−1.92	—	−1.92	—
	5	0.00	—	−1.00	—
	10	1.86	1.25	1.32	0.94
	15	4.31	2.63*	2.59	1.60
	20	3.80	2.38*	2.67	1.60
	25	7.54	4.19***	5.81	3.32**
	30	7.24	3.93***	4.65	2.72**
	35	7.34	3.74***	4.47	2.60*
	40	6.46	3.11***	4.87	2.40*

* Significant at the .05 level.
** Significant at the .01 level.
*** Significant at the .001 level.

[5] It is interesting to note that the application of the cumulative mean method to the results of another study reported by Kimble (8) does not change the theoretical significance of the results. It merely reduces the absolute magnitude of $_sI_R$ from 4.6 to 3.7.

(and the performance scores seem to bear out this assumption), the appearance of ${}_sI_R$ somewhat later in Kimble's groups than in the low motivation groups of the present experiment also fits the prediction. So likewise do the results of Kimble's more recent study if it is assumed that the man-to-man competitive situation which he utilized resulted in a higher level of motivation than that attained in either his earlier experiments with the alphabet printing task or in both motivation conditions of the present experiment. Using 5-min. practice periods with the pursuit rotor and 6-min. rest intervals, he failed to find evidence of conditioned inhibition. With high motivation, the *S*s were able to resist the resting tendency produced by I_R for the full 5 min. of work so that no resting occurred *during* learning and all the I_R could be dissipated during the 6-min. rest. Thus, the threshold of I_R necessary before ${}_sI_R$ can develop was never reached with this high degree of motivation.

The second prediction stated that late in learning the *rate* of development of ${}_sI_R$ should be less with high motivation than with low motivation. Reference to Table IV indicates that this prediction was also confirmed by the results. It may be seen that, regardless of the method of measurement, the increases in the amount of ${}_sI_R$ level off after 25 pre-rest trials with high motivation, while this starts to occur only after 30 pre-rest trials with low motivation. In general, these results agree with those of Kimble (7) who found that the asymptote had not yet been reached after 30 pre-rest trials.

The third prediction stated that late in learning the *amount* of ${}_sI_R$ should be less with high motivation than with low motivation. Again, the results presented in Tables IV and V clearly confirm this prediction. It may be seen that the difference between the two motivation conditions becomes greatest beyond 25 pre-rest trials. Further confirmation was gotten by an analysis of variance of the score differences which are presented in Table IV. The *F*-ratios for both motivation and amount of pre-rest practice were found to be significant beyond the .05 level of confidence regardless of the method of measuring ${}_sI_R$.

TABLE V

MEAN DIFFERENCE BETWEEN THE LOW AND HIGH MOTIVATION CONDITIONS IN THE AMOUNT BY WHICH THE SPACED GROUP EXCEEDS THE REMINISCENCE POINT

Group[a]	Kimble's Method		Cumulative Mean Method	
	Mean Diff.	t[b]	Mean Diff.	t[b]
5	2.55	7.97[d]	2.31	7.00[d]
10	−0.02[c]	0.07	−0.11[c]	0.34
15	2.04	6.80[d]	2.07	6.47[d]
20	3.59	11.22[d]	2.14	6.48[d]
25	0.05	0.16	−1.60[c]	4.85[d]
30	2.82	9.40[d]	3.06	9.56[d]
35	1.14	3.56[d]	2.32	6.63[d]
40	3.78	11.45[d]	2.27	6.49[d]

[a] Group 2 is omitted because the post-rest performance of this group in both motivation conditions was superior to the performance of the spaced group.
[b] These t-values are based upon the estimate of error variance obtained from an analysis of variance of the score differences presented in Table IV.
[c] The minus sign indicates that the value for the high motivation group was greater than that for the low motivation group.
[d] Significant at the .001 level of confidence.

SUMMARY

The alphabet printing task was administered to 538 *S*s under two conditions of motivation, low and high. Under each motivation condition a massed control group practiced under conditions of maximal inhibition, a spaced control group practiced under conditions of low inhibition, and nine experimental groups worked under

conditions of massed practice except for the introduction of a 10-min. rest after varying amounts of practice. The results and principal conclusions are as follows:

1. Spaced practice was found to be definitely superior to massed practice.
2. High motivation resulted in performance which was significantly superior to that with low motivation, the difference becoming progressively greater as practice continued.
3. Reactive inhibition (I_R) was built up extremely rapidly with both high and low motivation.
4. Less I_R was present late in learning with low motivation than with high motivation.
5. The decrease in reminiscence which usually occurs late in learning was clearly shown with low motivation, but was less marked with high motivation.
6. No conditioned inhibition ($_sI_R$) was present very early in learning under either motivation condition, and its initial appearance occurred later with high motivation than with low motivation.
7. Late in learning the amount and rate of development of $_sI_R$ were less with high motivation than with low motivation.
8. The results of this experiment are in substantial agreement with the hypothesis that high motivation should raise the critical (tolerable) level of reactive inhibition necessary to produce the automatic resting response.

(Manuscript received
September 21, 1950)

References

1. Ammons, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.
2. Ammons, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393–411.
3. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. Irion, A. L. Retention as a function of amount of pre-recall warming up. *Amer. Psychologist*, 1949, **4**, 219–220. (Abstract)
5. Irion, A. L. Reminiscence in pursuit rotor learning as a function of rest and amount of pre-rest practice. *J. exp. Psychol.*, 1949, **39**, 492–499.
6. Kientzle, M. J. Properties of learning curves under varied distributions of practice. *J. exp. Psychol.*, 1946, **36**, 187–211.
7. Kimble, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.
8. Kimble, G. A. Performance and reminiscence in motor learning as a function of the degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500–510.
9. Kimble, G. A. Evidence for the role of motivation in determining the amount of reminiscence in pursuit rotor learning. *J. exp. Psychol.*, 1950, **40**, 248–253.
10. Kimble, G. A., and Horenstein, B. R. Reminiscence in motor learning as a function of length of interpolated rest. *J. exp. Psychol.*, 1948, **38**, 239–244.
11. McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.
12. Ruch, F. L., and Warren, N. *Working with psychology.* Chicago: Scott Foresman, 1941.
13. Silver, R. J. The effect of the amount and distribution of warming-up activity on retention in motor learning. Unpublished Ph.D. thesis, New York Univ., 1951.

THE ACQUIRED DISTINCTIVENESS OF CUES: THE ROLE OF DISCRIMINATIVE VERBAL RESPONSES IN FACILITATING THE ACQUISITION OF DISCRIMINATIVE MOTOR RESPONSES

BY IRMA L. ROSSMAN AND ALBERT E. GOSS[1]

The University of Massachusetts

Miller (**24**), extending his earlier work with Dollard (**25**), has proposed that:

> According to stimulus-response theory, learning to respond with highly distinctive names to similar stimulus situations should tend to lessen the generalization of other responses from one of these situations to another since the stimuli produced by responding with the distinctive name will tend to increase the differences in the stimulus patterns of the two situations. Increased differentiation based on this mechanism has been called acquired distinctiveness of cues.
>
> On the other hand, if the individual learns to respond to two quite different situations with the same verbal response, the stimuli produced by this response will be a common element mediating an increased amount of generalization from one situation to the other. This has been called acquired equivalence of cues, or secondary generalization (**24**, p. 174).

Although experimental demonstrations of semantic generalization (**4**, **5**, **7**) and the studies of Shipley (**30**) and Lumsdaine (**22**) are indirectly relevant to the above analysis of the acquired equivalence of cues, in only one study (**2**) were the problem and experimental design directly formulated in Miller's terms. In this experiment Birge found that generalization or positive transfer of a manipulative response from one stimulus to another stimulus occurred more readily when young children had previously learned to give the same nonsense syllable response to the two stimuli for the manipulative response. These findings, supplemented by data which indicated that overt verbalization facilitated positive transfer, were consistent with expectations based on Miller's theoretical analysis.

Lawrence (**19**, **20**) and subsequently Jaynes (**18**) have reported data for discrimination learning in the rat which are consistent with predictions derived from Miller's analysis of the acquired distinctiveness of cues. However, at the time the present study was initiated there appeared to be no studies of acquired distinctiveness with human *Ss* which could have been considered directly relevant investigations of Miller's hypothesis. Accordingly, it was a major purpose of this study to investigate the hypothesis that *the acquisition of different verbal responses (names) to similar external stimuli will facilitate the subsequent acquisition of discriminative motor responses to the same external stimuli.*

Subsequent to the completion of the present study, Gagné and Baker (**8**) have reported that the amount of positive transfer to a motor discrimination task involving similar color-position cue combinations is a direct function of the amount of prior experience in distinguishing among these cues by means of different letter labels. Although not explicitly formulated in Miller's terms these results are in accordance with his analysis of acquired distinctiveness. In a later study Baker and Wylie (**1**) required three groups to learn a motor discrimination after the groups had had 0, 8, or 24 rehearsals in associating the names for the correct motor responses with the names for the color-position cues for the motor responses. While eight re-

[1] A more complete report may be found in (**27**). The first author was responsible for collecting and processing the data while the second author is in part responsible for the manuscript in its present form. Parts of these data were presented at the 1950 meetings of the American Psychological Association.

hearsal trials did not lead to significant positive transfer, 24 of such trials resulted in significant transfer effects. If the assumption that the distinctive verbal responses associated with the names for the color-position cues generalized to the actual cues during the motor discrimination trials is granted, these results are also consistent with the notion of the acquired distinctiveness of cues.

In the present investigation translation of the above hypothesis into experimental procedures involved comparison of the rates of acquisition of a motor discrimination by a group that had previously learned a verbal discrimination to a rigorous criterion, and groups that had been given a very small number of verbal learning trials.

Because the present study was considered to be preliminary to a series of studies stemming from Miller's analysis of the functional role of verbal cues two subsidiary and an additional major hypothesis were explored. In order to test one subsidiary hypothesis, that *overt verbalization during motor learning will lead to more rapid learning than covert verbalization*, as suggested by Birge's findings of the facilitative effects of overt verbalization, two groups which had learned the verbal discrimination to a criterion were used with different instructions during the motor learning.

Work reported by Gibson and Gagné and Foster as well as Miller's treatment of acquired distinctiveness of cues provided the theoretical background for a second subsidiary question. Gibson predicted (**11**) and subsequently verified (**13**) the existence of an initial increase and subsequent decrease in intra-list generalization during the course of paired-associate learning. In a subsequent investigation of transfer from practice on components in a discriminative motor task Gagné and Foster (**9**) also found evidence which was consistent with Gibson's hypothesis of an upward-downward trend in generalization. It is an implication of Miller's approach that increased generalization of verbal responses to stimuli should increase the difficulty of learning motor discrimination to the same stimuli. These considerations led to the hypothesis that, due to increased generalization of verbal responses, *a small number of verbal discrimination trials will lead to less rapid learning of a motor discrimination than either one or many verbal discrimination trials.* Two groups with a small number of verbal learning trials, specifically one and four trials, were used to investigate this hypothesis.

Within the framework of stimulus-response language Miller has interpreted the Freudian concept of primal repression as a decrement in the strength of verbal responses resulting from conflict between the verbal responses and anxiety-motivated responses of avoiding overt verbalization.[2] In the specific case of the mechanisms of acquired distinctiveness or equivalence of cues, the conflict would be between verbal responses which have functional significance in discriminative or generalizing behavior and anxiety-motivated avoidance responses. Behaviorally this conflict would be expressed directly as a decrement in the measured strength of verbal response. It would also be reflected indirectly as a decrement in the strength of discriminative or generalizing behavior which was functionally dependent on the verbal response-produced stimuli. It is the latter or indirect consequence of

[2] With the exception that covert or internal responses (anxiety) and anxiety-motivated overt responses have been distinguished, this conception of repression is essentially similar to Sears' (**28**) stimulus-response analysis of primal repression. Previous experimental work on repression has been summarized (**14, 28, 29, 32**).

repression which led to a second major hypothesis, that *repression of the verbal responses mediating acquired distinctiveness of cues will lower the rate of acquisition of discriminative motor responses.*

Apparatus and Stimulus Materials

Apparatus.—The stimulus materials were presented in the exposure slot of a memory drum at the rate of one every 2 sec. The interval between trials was 8 sec. Throughout the training period *E* was seated behind the drum out of the direct view of *S*. Additional apparatus consisted of a buzzer, a shocking device, and a motor-response device. The shocking device was an inductorium with a 1.5-v. dry cell current source. Shock was administered to *S* through a small wrist band with metal contacts which served as electrodes on the inside and outside surfaces of the wrist. The motor response device, a $6 \times 6 \times 1$ in. box, was constructed to permit the movement of a $3\frac{1}{2}$ in. lever along $5\frac{1}{4}$ in. slots which started from a point near the middle of one side of the top surface and extended diagonally forward toward the right and left corners of the opposite side. Movement of the lever to the end of the diagonal slots closed a circuit which in turn lighted one or the other of two lights arranged behind the drum to inform *E* of the direction of *S*'s response. The response device was placed in front of the right side of the shield of the drum; movement of the lever was diagonally left or right toward the shield.

Stimulus materials.—Lists of 16 and 12 paired-associate units and a list of 12 nonsense figures were employed as stimulus materials. The stimulus members of the 16-unit list consisted of eight of Gibson's (12) 13 standard figures and the eight corresponding figures of first-degree similarity. The response members of the units were eight pairs of Hull's (16) nonsense syllables of approximately equal (within ±4 per cent) and low (mean = 21.2 per cent) association values. These syllable pairs were selected so that the muscular movements required for saying one member of each pair would be antagonistic to the movements required for saying the other member. Specifically, the first and last sounds (consonants) of one member of each pair of syllables were stops while those of the other member were continuants (21). The eight stop syllables were assigned to the standard figures and the eight continuant syllables to the figures of first-degree similarity. Since the 16-unit list proved to be very difficult to learn, six of the original eight pairs of similar figures and their paired syllables were used to construct an easier 12-unit list. The same six pairs of figures without the paired syllables were then employed to make up a nonsense figure list.

Both figure-syllable pairs and figures alone were presented in four sequences in which each pair or figure appeared in a different quarter of the list during each sequence. Pairs containing similar figures or similar figures alone did not follow one another.

Since a maximum of eight paired-associate units could be presented with one rotation of the drum, it was necessary to divide the 16- and 12-unit lists into equal halves. The equal halves of each of the four sequences were placed in adjacent columns on the tape. The *S* shifted the exposure slots on the face of the drum when such was required. This required from 5 to 10 sec.

Experiment I: Effects of Shock

Purpose

This preliminary experiment was designed to determine whether shock would decrease the amount of retention of well-learned verbal responses. Demonstration of repression due to electric shock is an essential condition to the test of the hypothesis relating repression to the effects of acquired distinctiveness of cues.

Procedure

Acquisition.—Thirty-seven undergraduate students in psychology at the University of Massachusetts learned the 16-unit figure-syllable list to three consecutive errorless trials. Seven of these were eliminated for failure to reach the criterion in 45 trials.

The 30 remaining *S*s were divided into shock and no-shock groups of 15 *S*s which were matched by pairing each shock *S* with an individual no-shock *S* to achieve equal mean trials to learn (31.5) and standard deviations (7.5) with maximal matching in terms of sex. Individual shock and no-shock *S*s were paired by assigning each *S* who reached the criterion within ±2 trials of a no-shock *S*, preferably of like sex, to the shock condition. The assignment was made as the *S* completed the third criterion trial thus permitting the introduction of the shock trial without lengthening the usual 8-sec. inter-trial interval.

The *S*s were given standard instructions for

paired-associate learning with the following specific variants: the syllables were anticipated by pronouncing them, *S* was instructed to guess when unsure, and *S* was instructed to pronounce the correct syllable when it appeared, if an incorrect anticipation had been made.

Shock.—All *S*s had had their pain or discomfort thresholds for shock determined for their non-preferred hands before training. Those in the shock group were given one additional trial on the list after reaching the criterion, during which all anticipatory syllable responses whether correct or incorrect were followed by shock. If no response was made, the shock was administered after the syllable was exposed on the drum. The no-shock group received the same additional trial, but without shock.

Retention.—Eight seconds after the additional trial, with or without shock, both groups were given two additional trials. At the end of the retention test the *S*s of the shock group were given a short questionnaire designed to determine their reactions to the shock and retention trials.

Results

In Table I are shown the results of Exp. I. On the post-criterial trial during which shock was administered to the shock group, fewer correct responses were made. This difference is significant at better than the .01 level ($t = 4.23$). On the first and second retention trials, the shock group likewise made fewer correct responses than the no-shock group. However, the differences are not highly significant, even though consistent. For the first retention trial t is 2.40 ($.05 > p > .02$), for the second trial 1.78 ($p > .10$), and for both trials combined 2.23 ($.05 > p > .02$). In all estimates of significance the correlation between matched scores was taken into account ($df = 14$).

One *S* in the shock group failed to make any correct responses during the two retention trials. Elimination of this *S* and his paired *S* in the no-shock group increased all t-values somewhat and also increased the significance levels despite the loss of one degree of freedom.

Analysis of the questionnaire data indicates that the shock *S*s were irritated, puzzled, tense, and fearful, in that order of decreasing strength and frequency, during the shock trial. Tenseness and anxiety were the strongest reactions during the retention trials, with some *S*s reporting hesitancy to say the syllables.

Experiment II: Acquired Distinctiveness of Cues

Procedure

Experimental design.—The conditions of this experiment are outlined in Table II. The five groups were included to permit experimental answers to the following problems: (a) the determination of the facilitating effects of discriminative verbal responses by comparison of Group E-I with control Groups C-I and C-II; (b) the determination of the relative effectiveness of overt and covert verbalization by comparison of Groups E-I and E-III; (c) the determination of the retarding effects of shock by comparison of Groups E-I and E-II; and (d) the determination of the role of intra-list generalization by comparison of Groups C-I and C-II.

TABLE I

Effect of Shock for Correct or Incorrect Verbal Responses on Frequency of Correct Anticipations

Condition	Post-Criterial Trial		Retention Trials					
			First		Second		Both	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Shock	13.0	1.8	12.9	4.1	13.7	3.9	26.5	7.6
No-shock	15.2	0.7	15.5	0.8	15.6	0.6	31.1	1.0

TABLE II

CONDITIONS IN EXPERIMENT II

Groups (N = 15)	Verbal Learning	Post-Learning Training	Motor Learning
E-I	Mastery*	1 trial, no shock	20 trials
E-II	Mastery	1 trial, shock	20 trials
E-III	Mastery	1 trial, no shock	20 trials, overt verbalization
C-I	1 trial	None	20 trials
C-II	4 trials	None	20 trials

* Defined as three successive errorless trials.

Acquisition of verbal responses.—Sixty-three *Ss* to be assigned to Groups E-I, E-II, and E-III learned the 12-unit figure-syllable list to three consecutive errorless trials. Seventeen of these *Ss* failed to reach the criterion in 45 trials and were eliminated. Of the remaining 46, 45 were distributed into three equal groups according to the matching procedure employed in Exp. I, with the result that Groups E-I, E-II, and E-III required 24.0, 24.1, and 24.5 trials to learn with *SD*'s of 6.0, 6.0, and 5.5, respectively. An additional 15 *Ss* were assigned to Group C-I and given one learning trial on the 12-unit list. Similarly, 15 *Ss* were assigned to Group C-II and given four learning trials. Assignment to these groups was on a matching basis to be described in the section on the motor task.

Post-criterial trials.—The *Ss* in Group E-III were given shocks on one post-criterial trial. The interval between this trial and the third criterial trial was the usual 8 sec. As in Exp. I, Groups E-I and E-III were given one post-criterial trial without shock. All *Ss* had been previously tested to determine their pain or discomfort thresholds to shock in the non-preferred hand.

Acquisition of motor responses.—Approximately 1 min. after completion of the verbal learning all groups practiced a motor discrimination for 20 trials. Only the *Ss* of Group E-III were instructed to say the previously learned nonsense syllable response before or concomitant with the motor response. In the motor learning situation *Ss* learned to discriminate by means of left or right movements of the lever of the motor-response device, between pairs of similar nonsense figures. Specifically, they were to learn movement to the left as the correct response to each of the standard figures. Conversely, movement to the right was designated as the correct response to the corresponding figures of first-degree similarity. Knowledge of correct responses was given by means of a buzzer which was sounded by *E* after *S* pushed the lever to the end of the correct slot. No correction of incorrect responses was permitted.

The *Ss* in Groups C-I and C-II were paired with the *Ss* in Group E-I on the basis of the motor task. Each *S* of Group E-I was matched with *Ss* of Groups C-I and C-II who were within ±0.50 mean correct responses for the first block of 4 trials. Whenever possible *Ss* were further matched on the basis of sex, and *Ss* were run under the C-I and C-II conditions until the mean number of correct responses for the first four-trial blocks for the three groups were equal.

The instructions for the motor task merely told *S* to move the lever to the left or right when a figure appeared in the stimulus aperture and to return the lever to neutral after each response. They also informed *S* that a buzzer would sound if the direction of movement was correct, and that no correction of errors could be made. The *Ss* were not specifically informed that the figures to be seen were those to which nonsense syllable responses had previously been learned, except in the case of Group E-III. These *Ss* were instructed to make the previously learned nonsense-syllable response to a given figure before or concomitant with making the motor response to that figure.

Questionnaire.—All *Ss* were given questionnaires after the motor learning. One questionnaire, given to all groups, was designed to elicit additional information about the method used in the motor discrimination. The *Ss* of Group E-II filled out a second questionnaire on their reactions to the shock trial.

Results

With respect to the effect of shock on the verbal responses during the post-criterial trial, Table III clearly reveals the inhibitory effect of shock. The differences between the shock (E-II) and no-shock conditions (E-I,

TABLE III

EFFECT OF SHOCK ON FREQUENCY OF CORRECT RESPONSES DURING THE FIRST POST-CRITERIAL TRIAL

Group	Condition	Mean	*SD*
E-I	No-shock	12.0	0.0
E-II	Shock	9.4	1.4
E-III	No-shock	11.7	0.2

TABLE IV

MEAN CORRECT RESPONSES DURING MOTOR LEARNING

Group	Trials 1–4		Trials 5–20	
	Mean	*SD*	Mean	*SD*
E-I	6.02	0.6	31.0	4.0
E-II	5.88	0.7	28.1	4.5
E-III	5.55	1.0	27.2	5.0
C-I	6.02	0.6	26.0	2.6
C-II	6.02	0.7	25.3	3.4

E-III) are significant beyond the .01 level. The difference between Groups E-I and E-III is not significant. Group C-II made an average of 2.73 correct anticipations on the fourth verbal learning trial. Unfortunately, a mistake in the scoring procedure precluded the determination of intra-list generalization during this trial.

Motor learning data are presented in Table IV in terms of the means and standard deviations of correct responses made in Trials 1–4, the matching trials, and Trials 5–20, the comparison trials. In accordance with the pre-experimental criterion, Groups E-I, C-I, and C-II have equal first means for Trials 1–4. The mean values of 6.02 for these groups do not differ significantly from the somewhat lower means of Groups E-II and E-III.

Comparison of the mean for Trials 5–20 of Group E-I with those of Groups C-I and C-II yielded *t*'s of related measures of 3.28 ($p < .01$) and 3.32 ($p < .01$), respectively.[3] For corresponding differences between Groups E-I and E-II and Groups E-I and E-III, *t*'s of 1.68 ($.20 > p > .10$) and 2.40 ($.05 > p > .02$) were obtained. The difference between Groups C-I and C-II did not approach significance ($t = 0.56$). All *p*-values are based upon 14 *df*.

Learning curves on the motor discrimination are shown in Fig. 1. The measure of learning is the mean correct responses for successive four-trial units. Although Group E-I had reached a mean score of only 75 per cent correct responses by Trials 17–20, it is apparent that this group was consistently superior to the other groups. Group E-II was learning the motor discrimination at a somewhat lower rate than E-I. However, Group E-II was in turn performing at a higher level than the remaining groups during the last eight trials. The discrimination problem, as evidenced by the very slow rise of the curves, was much more difficult for Groups C-I, C-II, and E-III. However for trials 13–20 the performance of E-III was superior to that of the control groups.

Analysis of the questionnaire data concerning reactions to shock for Group E-II indicated that *S*s of that group felt puzzled, irritated, like hesitating to say the syllables, and tense, in order of decreasing strength and frequency. The groups with verbal learning experience used the syllables as facilitating devices to

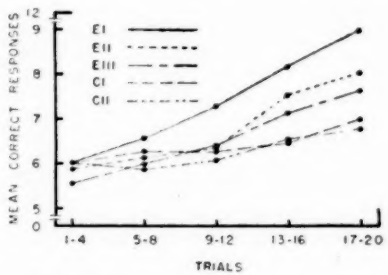

FIG. 1. Learning curves for mean correct responses for successive four-trial blocks

[3] Due to unconventional matching procedure it was considered inadvisable to utilize the more sophisticated techniques for group comparison in transfer studies advocated by Gagné, Foster, and Crowley (10).

some degree during the motor learning. However, they reported that looking for identifying parts was more frequently employed, and as subjectively evaluated, more helpful. The *Ss* of Groups C-I and C-II also reported that they used the identifying parts technique. There appeared to be only slight differences among the five groups in the frequency of use of this technique.

Discussion and Conclusions

Group E-I, which had mastered the verbal discrimination, learned the motor task more rapidly than Groups C-I and C-II, which had had only a very small number of verbal learning trials. Further, questionnaire results revealed that *Ss* of Group E-I had made covert nonsense-syllable responses during the course of motor learning and that, evaluated subjectively, these covert verbal responses had facilitated the acquisition of the motor discrimination. These results, which are consistent with those reported by Gagné and Baker (**8**), support the hypothesis that learning discriminative verbal responses facilitates the learning of discriminative motor responses to the same stimuli. However, before the superiority of Group E-I can be interpreted as an unequivocal confirmation of the specific hypothesis of transfer based upon the mechanism of the acquired distinctiveness of cues, the significance of the following conditions must be investigated.

First, since the questionnaire results revealed that "looking for and identifying parts" was the most frequently used, and, as subjectively evaluated, most facilitative technique for all groups, it is possible that the *Ss* of Group E-I learned the motor discrimination more rapidly because of their prior experience in using this technique. On closer analysis it would appear that the *Ss* could have interpreted the phrase "looking for and identifying parts" as referring to any one or all of three distinct techniques: (a) learning to select parts of pairs of similar figures which were more discriminable than the whole figure;[4] (b) learning to select parts of the figure pairs which may or may not have been more discriminable, and then learning to make the nonsense-syllable response to the selected parts; and (c) learning to select parts which may or may not have been more discriminable and then identifying the parts with familiar rather than nonsense-syllable names. As a consequence, the above information is of little value as the basis of alternative explanations of the results of the comparisons of Groups E-I with C-I and E-I with C-II. It should be noted, however, that both the second and third of the possible interpretations involve variations of the mechanism of acquired distinctiveness.

Second, during the learning of the verbal discrimination *Ss* of Group E-I may have also learned that there were pairs of similar figures. This knowledge, in the form "go left for one member of the pair and go right for the other member," may have facilitated motor learning for this group.

Finally, it might be hypothesized that Group E-I was aided by factors such as warm-up and knowledge of the orders of the figures.

It is suggested that until future experimentation isolates and indicates the functional relevance of the above conditions, the comparisons involving Group E-I are best interpreted as supporting the first major hypothesis that the acquisition of discriminative verbal responses will facilitate the acquisition of discriminative motor responses to the same external stimuli.

Comparison of mean correct responses and of learning curves for Groups E-I and E-III, the overt verbalization group, revealed that contrary to the first subsidiary hypothesis of the superiority of overt verbalization, the former group learned the motor task more rapidly. Reports volunteered by *Ss* at the conclusion of the motor discrimination

[4] The nonsense figure stimuli were not "perceptually simple" cues. Hence it is possible that orientation behavior (**31**) or "looking for and identifying" responses were relevant to the early stages of both verbal and motor discrimination learning.

provide the basis for a tentative explanation of the unexpectedly low performance of Group E-III. In general, *Ss* found that they did not have enough time to think of and to make the verbal response and to make the motor response during the 2-sec. response interval. The instructions to say the syllables aloud apparently introduced an interference factor which retarded the acquisition of the motor discrimination.

The second of the subsidiary hypotheses, that Group C-II, which had four verbal learning trials, would learn the motor discrimination less rapidly than Group C-I, the one-trial group, was based on Gibson's (**13**) findings that intra-list generalization increased to a maximum during the third and fourth tenths of total learning trials. Comparison of total scores revealed a slight but not statistically significant difference in favor of Group C-I. In part because of the previously noted mistake in scoring which precluded assessing the degree of intra-list generalization for Group C-II, the data do not permit the isolation of the factors which may have accounted for the failure to obtain a significantly retarded performance for Group C-II.

Under the shock-trial conditions of Exp. I, according to principles advanced by Mowrer (**26**), Miller (**23**), and others (**3, 6**), internal and overt responses to shock should have been conditioned to the cue pattern of figure plus verbal response-produced stimulation. This conditioned form of the internal response to pain has been termed anxiety or fear. On the subsequent retention trials, presentation of the figure alone should have (a) by the principle of generalization (**17**), elicited the internal (anxiety, fear) and overt responses previously aroused by shock, and (b) reinstated the cue-pattern of figure plus response-produced stimulation most strongly conditioned to anxiety. However, the increase in anxiety occasioned by reinstating the cue-pattern could have been prevented by failing to respond. Since shock was an immediate consequence of both correct and incorrect responses but was given only after exposure of the correct syllable in the case of failure to anticipate, avoidance of the verbal responses would have had the two-fold effect of preventing an increase in anxiety and of postponing shock. It was anticipated, therefore, that as a result of the presence of conflicting anxiety-motivated avoidance responses the strength of the verbal responses on the two post-shock retention trials would be reduced.[5] This expectation was tentatively confirmed by small decrements, significant at better than the .05 level, in the first and combined retention test scores of the shock group.

These results provided the basis for the second major hypothesis of Exp. II, that due to repression of the discriminative verbal responses the shock group (E-II) would learn the motor task less rapidly than Group E-I. While consistent with this hypothesis, the observed superiority of Group E-I in mean correct responses was not statistically significant.

The failure to obtain a significant

[5] Farber (**6**) has hypothesized that responses which have been followed by shock will be strengthened relative to other responses under conditions of unavoidable punishment and/or high initial strength of the punished response. Further, Gwinn (**15**) has demonstrated that if the punished response is compatible with the escape response, punishment will, at least initially, strengthen the punished response. It should be noted, therefore, that the predicted decrement in verbal response strength is restricted to the one shock trial of the present experiments. It is possible that continued shock trials might lead ultimately to increased verbal response strength.

difference between Groups E-I and E-II may have been due to failure to achieve a fairly large and stable repression with only one shock trial. In this case, the present experiment did not realize the conditions necessary for an adequate test of Miller's repression hypothesis. Further, while it is doubtful that shortening the figure-syllable list by four units would have significantly reduced the percentage decrement in response strength attributable to one shock trial, this possibility is not demonstrably irrelevant.

Summary

In this study several hypotheses stemming from the concept of the acquired distinctiveness of cues were investigated. Stimulus materials were a 12-unit figure-syllable paired-associate list and a list of the same 12 figures alone, both of which were presented at a 2-sec. rate on a memory drum. Additional apparatus consisted of motor response and shocking devices.

Three matched groups of 15 *S*s each were used to determine whether the acquisition of discriminative verbal responses facilitates the acquisition of discriminative motor responses to the same stimuli. The results indicated that a group (E-I) which had learned to mastery the discrimination between pairs of similar figures by means of nonsense syllable responses acquired discriminative motor responses to the same figure stimuli significantly more rapidly than groups (C-I and C-II) which had been given only one and four verbal learning trials respectively.

The motor learning performance of Group E-I was superior, although not significantly so, to the performance of another group (E-III) which had also learned the verbal discrimination to mastery and was then instructed to verbalize verbal responses overtly during the motor learning trials. The hypothesis that overt verbalization of discriminative verbal responses would have greater facilitative effects than covert verbalization was not supported, but the short time (2 sec.) allowed for the verbal and motor responses may have been an important factor in this result.

Since the difference between Groups C-I and C-II was not significant, the hypothesis that, due to intra-list generalization, the group with only one verbal learning trial (C-I) would learn more rapidly than the group with four verbal learning trials (C-II), was not supported.

A preliminary investigation of the effects of shock for correct or incorrect responses after mastery of a verbal discrimination suggested that shock occasioned a reduction in the immediate retention of the verbal responses. It was possible, therefore, to investigate the hypothesis that repression of the verbal responses mediating acquired distinctiveness of cues by means of electric shock would lower the rate of acquisition of discriminative motor responses. Although the direction of differences obtained was in the predicted direction, they were not statistically significant.

(Manuscript received
September 16, 1950)

References

1. Baker, Katherine E., and Wylie, Ruth C. Transfer of verbal training to a motor task. *J. exp. Psychol.*, 1950, **40**, 632–638.
2. Birge, J. S. The role of verbal responses in transfer. Unpublished Ph.D. dissertation, Yale Univ., 1941.
3. Brown, J. S., and Jacobs, A. The role of fear in the motivation and acquisition

of responses. *J. exp. Psychol.*, 1949, **39**, 747–759.

4. Cofer, C. N., and Foley, J. P., Jr. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, 1942, **49**, 513–540.
5. Cofer, C. N., Janis, M. G., and Rowell, M. M. Mediated generalization and the interpretation of verbal behavior: III. Experimental study of antonym gradients. *J. exp. Psychol.*, 1943, **43**, 266–269.
6. Farber, I. E. Response fixation under anxiety and non-anxiety conditions. *J. exp. Psychol.*, 1948, **38**, 111–131.
7. Foley, J. P., Jr., and Cofer, C. N. Mediated generalization and the interpretation of verbal behavior: II. Experimental study of certain homophone and synonym gradients. *J. exp. Psychol.*, 1943, **32**, 168–175.
8. Gagné, R. M., and Baker, Katherine E. Stimulus pre-differentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439–451.
9. Gagné, R. M., and Foster, Harriett. Transfer of training from practice on components in a motor skill. *J. exp. Psychol.*, 1949, **39**, 47–68.
10. Gagné, R. M., Foster, Harriett, and Crowley, M. E. The measurement of transfer of training. *Psychol. Bull.*, 1948, **45**, 97–130.
11. Gibson, Eleanor J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
12. Gibson, Eleanor J. Retroactive inhibition as a function of the degree of generalization between tasks. *J. exp. Psychol.*, 1941, **28**, 93–115.
13. Gibson, Eleanor J. Intra-list generalization as a factor in verbal learning. *J. exp. Psychol.*, 1942, **30**, 185–200.
14. Glixman, A. F. An analysis of the use of the interruption-technique in experimental studies of "repression." *Psychol. Bull.*, 1948, **45**, 491–506.
15. Gwinn, G. T. Effects of punishment upon acts motivated by fear. *J. exp. Psychol.*, 1949, **39**, 260–269.
16. Hull, C. L. The meaningfulness of 320 selected nonsense syllables. *Amer. J. Psychol.*, 1933, **45**, 730–734.
17. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
18. Jaynes, J. Learning a second response to a cue as a function of the magnitude of the first. *J. comp. physiol. Psychol.*, 1950, **43**, 398–408.
19. Lawrence, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–784.
20. Lawrence, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40**, 175–188.
21. Lewis, C. L. *American speech.* Chicago: Scott, Foresman, 1916.
22. Lumsdaine, A. A. Conditioned eyelid responses as mediating generalized conditioned finger reactions. *Psychol. Bull.*, 1939, **36**, 650. (Abstract)
23. Miller, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89–101.
24. Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.
25. Miller, N. E., and Dollard, J. *Social learning and imitation.* New Haven: Yale Univ. Press, 1941.
26. Mowrer, O. H. A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.*, 1939, **46**, 553–565.
27. Rossman, Irma L. The acquired distinctiveness of cues: the role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. Unpublished M.A. thesis, Univ. Massachusetts, 1950.
28. Sears, R. R. Functional abnormalities of memory with special reference to amnesia. *Psychol. Bull.*, 1936, **33**, 229–274.
29. Sears, R. R. *Survey of objective studies of psychoanalytic concepts.* New York: Soc. Sci. Res. Council, 1943, Bull. 51.
30. Shipley, W. C. Indirect conditioning. *J. gen. Psychol.*, 1935, **12**, 337–357.
31. Spence, K. W. Continuous vs. non-continuous interpretations of discrimination learning. *Psychol. Rev.*, 1940, **47**, 271–288.
32. Zeller, A. F. An experimental analogue of repression. *Psychol. Bull.*, 1950, **47**, 39–51.

ANXIETY AND STRENGTH OF THE UCS AS DETERMINERS OF THE AMOUNT OF EYELID CONDITIONING

BY K. W. SPENCE AND JANET TAYLOR [1]

State University of Iowa

In a recent experiment by one of the authors (**9**), an attempt was made to evaluate the hypothesis that the amount of conditioning exhibited in a defense conditioning situation is a direct function of the level of total effective drive. Following Hull (**3**), it was assumed that the value of drive (*D*) operating at the time of occurrence of a conditioned response is determined not only by the relevant need (i.e., the one which is reduced by the response under consideration), but also by the aggregate strength of all other primary and secondary needs operative at the moment. The chief implications of this assumption are supported by concrete experimental evidence (**1, 5, 10**).

On the assumption that variations in what is described psychiatrically as "manifest anxiety" reflect variations in generalized drive levels, groups of anxious and non-anxious *Ss* chosen on the basis of extreme scores on a test of manifest anxiety,[2] were compared with respect to the development of the conditioned eyelid response. The results indicated that the anxious group was consistently superior in the amount of conditioning exhibited throughout the course of conditioning trials, the differences between the two groups being highly significant statistically (**9**).

The present study represents a further investigation of the conditioning of anxious and non-anxious *Ss*. The experiment was designed to provide data on the effects of varying degrees of intensity of the unconditioned stimulus (UCS) on the amount of conditioning exhibited by anxious and non-anxious *Ss*, chosen independently in the same manner as the previous study.

A study by Passey (**7**) constitutes the only evidence now available in the literature concerning the effect of variation of the intensity of the air puff on the development of the conditioned eyelid response. Four groups of ten *Ss* were given 50 paired presentations of a tone and an air puff, each group receiving a different intensity of puff as the UCS. The results indicated that the amount of conditioning was a positive function of the intensity of the UCS. According to Passey's analysis, the rate of acquisition exhibited a direct relationship to the logarithm of the intensity. Unfortunately, none of the groups appeared to have reached the limits of conditioning by the end of the training trials so that a comparison of the asymptotes of the groups was not possible.

In an as yet unpublished study by the present writers (**8**) a comparison was made of conditioning with an air puff resulting from a 100-mm. fall of a column of mercury with a puff resulting from a 30-mm. fall. The curve

[1] Now at Northwestern University.

[2] The manifest anxiety scale consisted of 65 items from the Minnesota Multiphasic Personality Inventory judged by clinicians to be indicative of manifest anxiety. Selection of *Ss* was made by taking the upper and lower 15 per cent of a distribution of scores of 352 students in introductory psychology. A more complete description of the test is available elsewhere (**9**).

for the group with the strong puff rose more rapidly to a higher asymptote than that for the weak puff group. The differences between the total number of CR's for all trials were found to be significant between the .01 and .05 levels of confidence.

Procedure

Subjects.—A test of manifest anxiety was administered in group form to 574 students in a course in introductory psychology. The measures, based on 50 items, ranged from a low anxiety score of zero to a high anxiety score of 41[1]. The selection of *Ss* for the anxious and non-anxious groups was made by taking the individuals who scored in the upper 20 per cent (score of 24 or above) and the lower 20 per cent (score of 9 or below) of the distribution. One hundred and twenty-one *Ss* were run in the experiment, of which eight were lost either because of apparatus difficulties or *E*'s error, and two because of excessive spontaneous blinking. Of the remaining 111 *Ss*, 11 were discarded, five from the anxious groups and six from the non-anxious groups because they did not meet certain latency criteria that were established as a means of eliminating *Ss* who were believed to be making voluntary eyeblinks to the conditioned stimulus. Quite in contrast to the usual situation that has held in conditioning studies in the Iowa Laboratory, it was found that a high percentage (82 per cent) of the *Ss* recognized and were able to state the purpose of the experiment, i.e., as concerned with eyelid conditioning. This contrasts very sharply with the percentage of 23 found by Taylor (9) and approximately 20 per cent found in another study (8).

The reason for the unusual degree of sophistication on the part of the *Ss* was easy to trace. It resulted from a lecture given just prior to the start of the experiment on the CR in which details of the method of recording eyelid CR's were described by slides. From verbal reports of the *Ss* it soon became apparent that many of them sought to cooperate and facilitate the conditioning by voluntarily closing their eye to the conditioned light signal and keeping it closed until after the occurrence of the UCS. The pattern of such responses was quite characteristic in form and could be easily differentiated. They also tended to be extremely short in latency. In order to eliminate such *Ss* by an objective criterion the latencies of all *Ss* were examined and the percentage of CR's less than 300 msec. in latency was computed for each *S*. The median of this value was found to be 14.3 per cent. There was a definite bimodality to distribution with the distribution dropping to zero frequency for percentage scores between 40 per cent and 50 per cent with a second mode appearing above this value. Accordingly, it was decided to exclude the data of all *Ss* who gave more than 50 per cent CR's with latencies less than 300 msec.

Fifty of the 100 remaining *Ss* were from the high (anxious) end of the distribution of anxiety scores and 50 from the low (non-anxious) end. Half of each of these two groups was given a relatively strong puff of air and half a weak puff. There were thus four sub-groups of 25 *Ss* each, designated as follows: Anxious strong, anxious weak, non-anxious strong and non-anxious weak.

Apparatus and method of recording.—The *S* was seated in a dental chair in a semi-darkened room adjoining that in which the recording apparatus and stimulus controls were located. Fixed in a reclining position in the chair with his head braced against a head rest, *S* was instructed to blink on receiving a ready signal, and then to fixate on a 6 cm. circular milk glass window or disc, placed at a distance of 132 cm. in front of him. The brightness of the milk glass disc between trials was .05 apparent foot candles.

The CS consisted in an increase in brightness of the disc to 1.51 apparent foot candles. The duration of the CS on each trial was 550 msec. The USC, a puff of air, followed the onset of the CS by 520 msec. and was delivered to the external canthus of the right eye. The air puff was produced by the fall of a column of mercury, magnetically released, in a manometer. The fall was 100 mm. for the strong UCS, and 30 mm. for the weak UCS.

The onset of the CS was controlled by the closing of a switch on an electronic timer. The latter controlled the duration of the CS and the onset of the UCS at the appropriate time interval. The onset of the stimuli was recorded on a Brush polygraph by means of a Brush BL-902 pen motor.

The movement of the eyelid was recorded by a combination of mechanical and electrical means that provided a reasonably linear record of the position of the eyelid at all times. The method is being described in detail for publication elsewhere (4). Briefly, it involved the use of (a) a very lightweight microtorque potentiometer, (b) a fixed frequency oscillator, (c) an audio-frequency amplifier, (d) a full wave rectifier for rectifying the amplified signal, (e) a Brush BL-902 pen recorder and (f) a Brush polygraph.

The microtorque potentiometer was mounted on an adjustable head band just above *S*'s eye.

[1] On the basis of an item analysis of the 65 items in the original scale, the 50 most discriminating items were selected for the new scale.

A very light lever was attached at right angles to the variable (rotating) arm of the potentiometer and a mechanical linkage was accomplished between it and a tiny plastic lever fastened to the upper eyelid by adhesive tape. A small restoring spring was added to the potentiometer arm thus permitting it to follow the eyelid's motion. The system was very light and involved a minimum of work. Movement of the eyelid, with its accompanying rotation of the potentiometer arm, altered the voltage from an oscillator source applied to an amplifier. The output of the amplifier was then rectified and made to operate the Brush pen motor.

Conditioning procedure.—Each *S* first received three presentations of the light alone in order to check for any initial conditioned response tendency to the light. A single puff of air, unaccompanied by the light, was then administered and spontaneous blinking recorded for 40 sec. following onset of the puff. Immediately following the preliminary trials, all *S*s were given 100 paired presentations of the light and air puff. Inter-trial intervals of 15, 20, or 25 sec., averaging 20 sec. and arranged according to a fixed schedule, were used. At the end of the session, all *S*s were questioned as to their understanding of the purpose of the experiment and warned not to discuss the experiment with other members of the class.

Results and Discussion

Anxiety and conditioning.—Curves of acquisition of the CR, based on the median per cent of anticipatory responses occurring in successive blocks of ten trials for the four subgroups are presented in Fig. 1. An inspection of the curves reveals that the anxious groups (anxious-strong and anxious-weak) tend consistently to be above the non-anxious groups.

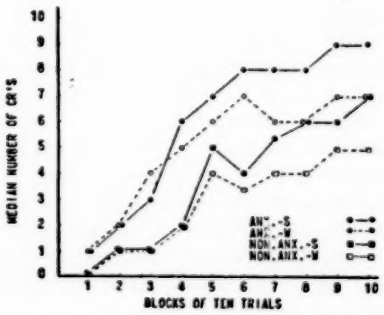

Fig. 1. Acquisition curves showing the median number of CR's for the four subgroups in successive blocks of ten trials

TABLE I

Mean Per Cent of CR's Exhibited on 100 Conditioning Trials

Strength of UCS	Anxious *S*'s		Non-Anxious *S*s		Total
	Mean	σ_M	Mean	σ_M	Mean
Strong	55.0	4.71	41.7	5.01	48.35
Weak	48.2	4.63	33.8	4.45	41.00
Total	51.60		37.75		

The total amount of conditioning in 100 trials exhibited by the four groups, shown in Table I, further indicates the superiority of the anxious groups. The results of an analysis of variance, used to determine the significance of the obtained differences, are summarized in Table II. The *F* for the anxious—non-anxious condition was found to be 8.59, which is significant at above the .01 level of confidence. In order to test the significance of the difference between the corresponding anxious and non-anxious groups, the *t*-test was employed. For the difference between the anxious-strong and non-anxious strong groups the *t* was 1.98, and that between the anxious and non-anxious weak groups was 2.14, both of which

TABLE II

Analysis of Variance for Trials 1–100

Source	*df*	Mean Square	*F**
Anxiety	1	4719.69	8.59
Strength of puff	1	1332.25	2.42
A × S	1	8.41	
Within groups	96	549.57	
Total	99		

* .05 level, $F = 3.94$; at .01 level, $F = 6.90$.

are significant at the .05 level of confidence.[4]

The superior conditioning of the anxious groups confirms the results of the previous study (9). Although the strength of puff used in the original investigation was somewhat less intense than that employed with the present strong puff groups, it is possible to make a rough comparison of the conditioning of the strong puff groups in the present study and those in the previous investigation.

A comparison of the two sets of non-anxious *Ss* reveals that the present non-anxious group exhibited a greater amount of total conditioning (39.7 vs. 27.9 per cent CR's in 80 trials). While it is reasonable to expect the non-anxious strong *Ss* to exhibit a slightly higher amount of conditioning because of the more intense air puff, it is possible to suggest a further factor that may have contributed to such a difference. As was mentioned previously, a large number of *Ss* (82 per cent) in the present investigation were aware of the purpose of the experiment and appeared to adopt a positive set that facilitated the conditioning. Although an attempt was made to eliminate *Ss* who exhibited voluntary responses to the light, it is possible that the presence of this facilitatory set in many of the *Ss* resulted in a higher amount of conditioning than would be found in more naïve *Ss*. Ample experimental evidence indicates the influence that set may have on the development of the CR and tends to support this suggestion (2, 6).

With respect to the anxious *Ss*, however, the anxious-strong group of the present investigation exhibited considerably *less* conditioning than the anxious *Ss* in the first study, in spite of the greater intensity of puff (39.96 vs. 59.60 per cent CR's in 80 trials). However, it should be noted that the criterion for inclusion in the anxiety group in the original experiment was a percentile rank of 89 or above on the anxiety scale, as compared to a percentile rank of only 80 in the present investigation. In order to check the possibility that the discrepancy in the range of anxiety scores in the two studies may have been responsible for the differences in the conditioning of the two sets of anxiety *Ss*, the conditioning scores for only those anxious-strong *Ss* with percentile ranks of 89 or better on the anxiety scale were computed. The amount of conditioning exhibited by this restricted anxiety-strong group was quite comparable to that of the anxious *Ss* in the original study, being 60.3 per cent for 80 trials.

[4] The results were also analyzed by the Mann-Whitney *U*-test. The differences between the anxious and non-anxious groups were found to be between the .02 and .05 levels of significance.

Strength of UCS and conditioning.— It is clear from an examination of the acquisition curves presented in Fig. 1 that the anxious and non-anxious strong puff groups tended to condition somewhat better then did the corresponding weak puff groups, particularly in the latter half of the training period. This superiority of the strong *Ss* is further revealed in the total amount of conditioning of each group, shown in Table I. However, the F obtained from the analysis of variance was 2.42, a value which does not reach significance at the .05 level of confidence.

The failure to find a significant difference between the strong and weak puff conditions contrasts sharply with another study (8) the writers have conducted in which the groups were selected at random from an introductory psychology class rather than by means of anxiety scores. In this investigation, in which the strengths of puff employed were the same as those utilized in the present experiment, the strong puff group exhibited markedly superior conditioning, especially in the first 50 trials. The results of Passey's study (7) also clearly demonstrated that strength of the unconditioned puff stimulus is a factor determining amount of conditioning.

One possible explanation of the failure to find significant differences in the present experiment is that this variable (strength of UCS) is not effective in the case of *Ss* at the extreme ends of the anxiety scale. In other words, the differences in strength of the UCS produce no marked difference in the level of drive (anxiety) of extremely anxious and extremely non-anxious subjects.

Another difference between the present study and that concerned with the effects of puff strength (8) was the

factor, already mentioned, of degree of sophistication of the *Ss*. The presence of a large number of sophisticated *Ss* in the present experiments may have obscured the differential effects of the strength of puff on the development of the CR.

Theoretical Interpretation

One possible interpretation of the difference in amount of conditioning exhibited by the different groups of *Ss* is that these groups, in Hull's terms, develop habit strength ($_sH_R$) at different rates and reach a different level of maximum habit strength. This is the suggestion that Passey (**7**) made concerning the effects of increased intensities of puff that he obtained. It should be noted, however, that according to Hull's theoretical formulation, the response measure is a function of both habit and drive strength: $R = f(H \times D)$. Thus, it is theoretically possible for the habit growth curves of the different groups of *Ss* to be the same and yet have their response curves increase at different rates to different asymptotes because of different drive (D) values.

Whether the different response measures obtained for the different groups in Passey's experiment and in the present study reflect differences in (a) habit strength, (b) drive strength or (c) both of these intervening variables, cannot be answered by the type of experimental design employed. In a further experiment conducted by the present writers (**8**) a specific attempt is being made to obtain evidence as to whether differences in habit strength are produced by different intensities of the UCS.

Summary

The experiment was designed to provide data on the effects of strong and weak intensities of the UCS on the amount of conditioning exhibited by anxious and non-anxious *Ss*, selected independently of the experimental situation.

Two groups of 50 *Ss* were chosen on the basis of extreme scores made on a test of manifest anxiety. The groups were run in a conditioned eyelid situation. Half of each group received a relatively strong puff of air as the UCS, the remaining half of each group a relatively weak puff.

Analysis of the results indicated that the anxious subgroups were consistently superior to the corresponding non-anxious groups in amount of conditioning throughout the course of the conditioning trials, the difference between the groups being significant statistically. Comparison of the strong- and weak-puff groups revealed that although the strong group exhibited a larger number of CR's throughout the conditioning, the differences from the corresponding weak groups were not significant. The results are discussed in terms of Hull's constructs of habit ($_sH_R$) and drive strength (D).

(Manuscript received October 2, 1950)

References

1. Amsel, A. The combination of a primary appetitional need with primary and secondary emotionally derived needs. *J. exp. Psychol.*, 1950, **40**, 1–14.
2. Hilgard, E. R., and Humphreys, L. G. The effect of supporting and antagonistic voluntary instructions on conditioned discrimination. *J. exp. Psychol.*, 1938, **22**, 291–304.
3. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. Hunter, T. A., Harker, G., and Spence, K. W. A new eyelid recording apparatus. (In preparation.)
5. Kendler, H. H. Drive interaction: I. Learning as a function of the simultaneous presence of hunger and thirst drives. *J. exp. Psychol.*, 1945, **35**, 96–109.

6. MILLER, J. M. The effect of inhibitory and facilitatory attitudes on eyelid conditioning. *Psychol. Bull.*, 1939, **36**, 577–578. (Abstract.)
7. PASSEY, G. E. The influence of the intensity of unconditioned stimulus upon acquisition of a conditioned response. *J. exp. Psychol.*, 1948, **38**, 420–428.
8. SPENCE, K. W., AND TAYLOR, JANET A. The effect of intensity of the unconditioned stimulus on the conditioned eyelid response. (In preparation.)
9. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
10. WEBB, W. B. The motivational aspect of an irrelevant drive in the behavior of the white rat. *J. exp. Psychol.*, 1949, **39**, 1–14.

MEMORY FOR PLEASANT AND UNPLEASANT EXPERIENCES: SOME METHODOLOGICAL CONSIDERATIONS

BY RALPH H. TURNER AND JOHN A. BARLOW

Oberlin College

The early work in this field was characterized by findings which contained so many inconsistencies that it became apparent that many of the factors relevant to the recall of pleasant and unpleasant experiences were not being adequately controlled. More recent experimentation has had the merit of incorporating the findings of previous work, with the result that today's findings contain fewer ambiguities. In this paper, the authors present the results of an experiment which points to two further respects in which inadequate methodology has characterized work in this field.

There are many respects in which experiments in this area have been inadequately controlled. (a) Early investigators attempted to use the very unreliable questionnaire method in which *S*s were asked whether they recalled pleasant or unpleasant experiences more readily. (b) In some instances as few as 15 *S*s have been used in the experimental work. (c) Words have often been selected by *E* as pleasant and unpleasant without determining whether the *S*s perceived them to be pleasant and unpleasant as assumed. (d) Some procedures have involved presenting *S*s with material to be learned, the *S*s later being asked to designate which material was pleasant and which was unpleasant. Such a procedure runs the risk that *S* will tend to designate as pleasant the material which was more readily learned. (e) Artificial laboratory situations, involving pleasant or unpleasant odors or colors or other sensations, have been used in inves igating the retention of P and U experiences. (f) Many *E*s who saw the inadequacy of using pleasant or unpleasant odors or colors in artificial laboratory experiences nevertheless used words derived from Thorndike's word list or some other source. (g) Much work has confined itself to a simple rating of the recalled material as to pleasantness or unpleasantness without a rating of the intensity accompanying such feeling. (h) Many experiments, such as those of Kowalewski (6) and Wohlgemuth (13), have asked for the recall of events following a vacation period when the affective tone of recent experiences is more likely to be judged to be pleasant. (i) Some work has been done on the assumption that the recall of a larger number of P than U words is sufficient to justify the conclusion that P experiences are remembered better than U experiences, but the work of Flügel (3) has shown the probability that more experiences judged to be P are available for recall. (j) Conclusions have often been drawn on the basis of one immediate recall, but Gilbert (4) has noted that delayed recall periods give opportunity for the relevant factors under investigation to operate and thus yield results more likely to correspond with the delayed recall found in everyday life situations. (k) Some *E*s have failed to note the fact that many experiences recalled during the second or third recall period were "new experiences" in the sense that they were not reported during the first recall period. This tendency to lump together the newly reported with previously reported experiences has obscured consistencies in the data which appeared when the data did not contain these contaminations. (l) The age of the *S*s has been overlooked as a relevant variable by most investigators with the result that there have been uncritical comparisons made between data derived from experiments using persons of widely divergent age groupings. (m) Intensity judgments have been made separately within the P and U categories on the assumption that corresponding points on the P and U intensity scales represent equivalent intensities. The study of P and U experiences with presumably equivalent intensities has thus led to the assumption that reported retention differentials could be attributed only to the q ıalitative differences inherent in P and U experiences. (n) Temporal sequence in the recall phase of the experiment is a factor which has not been given adequate attention in experimental methodology. Since the present experiment was designed to throw light on the problems raised by the last two criticisms, both need elaboration.

The Intensity Variable

Jersild (**5**), Sharp (**9**), O'Kelly and Steckle (**8**), Stagner (**10**), and Thomson (**11**) made no systematic study of intensity. Barrett (**1**) had *Ss* rate P and U words separately on a scale of one to three and concluded that the mean intensity of P and U words was the same because the mean scale rating of P and U words was not significantly different. This procedure is hardly justified since there is no way of knowing that the subjective one-to-three intensity scale for P experiences was distributed along the same intensity steps as the one-to-three scale used for judging U experiences. Cason (**2**) had his *Ss* rate experiences on a scale from extremely pleasant through indifferent to extremely unpleasant. With such a procedure there is no way of determining the extent to which an intensity on the P scale could be considered the equivalent of the corresponding point on the U intensity scale. Menzies' (**7**) work is subject to the same criticism since he had five intensity steps for pleasant and five minus steps for unpleasant experiences but had no way of determining intensity equivalents for his P and U scales. In the absence of knowledge to the contrary, the literal equivalence of corresponding points on intensity scales has often been assumed.

Perhaps the best handling of the intensity variable has been found in the work of Waters and Leeper (**12**). They used what amounted to a three-point intensity scale for P and for U experiences. By means of a refining procedure, which involved the utilization of only those experiences recalled the first and the second time with the same intensity rating, they obtained results free from the ambiguities characteristic of much previous work. Even this refinement, however, does not assure the comparability of corresponding points on P and U intensity scales. The fact that *Ss* were consistent from one time to the next in their intensity judgments of words points only to the reliability of judgments made separately within the P intensity scale and within the U intensity scale without in any way assuring us of comparability of corresponding points on these scales. A possible alternative to these methods of handling the intensity variable is detailed in the section on "procedure."

The Temporal Sequence Variable

Previous experiments have recognized the importance of temporal sequence in the *learning* phases by alternating P and U experiences. The same cannot be said of the *recall* phases of the experiments. In the classical experiment of Jersild (**5**), as well as in the experiments reported by Kowalewski (**6**) and Wohlgemuth (**13**), this factor was ignored as a variable. In the experiment of O'Kelly and Steckle (**8**) this variable is apparently recognized but the data are not analyzed to determine its importance. Moreover they do not report the absolute or relative size of the groups in which counterbalanced order was used so that it cannot be determined whether the factor of temporal sequence has been balanced out by pooling the results derived from the two groups.

Stagner (**10**) asked his *Ss* to list their P and U experiences on separate sheets but gave no indication as to which was requested first or whether any consistent plan was followed. Young (**14**) asked his *Ss* to put "pleasing" words in one column and "displeasing" in another. For three groups the pleasing words were to be listed in the left column, and in one

group the situation was reversed and the words "pleasing" and "displeasing" were interchanged in reading the directions. But since the *S*s were permitted to recall words in whatever temporal sequence they chose, there is no way of knowing whether the effects of a temporal sequence were balanced out. Unless there are specific directions to the contrary, there is a tendency to recall pleasant words first and then unpleasant, probably because the phrase "pleasant-unpleasant" suggests this sequence of operations. The fact that most *E*s have followed this "P then U" sequence, testifies to the validity of this assumption. The following section on "procedure" describes a method of investigating the importance of the temporal sequence variable.

Procedure

The procedures followed in conducting this experiment will be given fully in order that the results may be capable of interpretation by those who choose to define terms operationally. A total of 250 students taking introductory psychology served as *S*s. In order to study the intensity variable and the temporal sequence variable, these *S*s were divided into three groups. Table I indicates the procedures and the number of *S*s in each group.

TABLE I

Experimental Conditions

Group	Instructions	No. of Subjects		
		Listing P First	Listing U First	Total
A	Rank P experiences in intensity from most P as No. 1 to least P as *N*. Same with U experiences.	34	23	57
B	Same procedure as for Group A.	67	59	126
C	Rank all experiences, whether P or U, on *one scale*, from most intense as No. 1 to least intense as *N*.	29	38	67

After distributing papers bearing identifying numbers, the instructions to the *S*s at the time of the first (original) recall were given as follows:

The *E* said to all groups: "For a few minutes this morning we are going to conduct a survey of your recent emotionally tinged experiences. You will be given 4 min. in which to record all of the most (pleasant or unpleasant in accordance with Table I) experiences you have had during the past week. It is not necessary for you to write out a complete description of each incident but only a sentence or perhaps a word or phrase which will permit you to identify the experience later. Be as frank and as complete as possible in the time allotted. As your names are not written on these papers, your responses are completely anonymous."

After a pause of 4 min., *E* said to Groups A and B only: "Now please rank these experiences in intensity from most (unpleasant or pleasant) as number one to the least (unpleasant or pleasant)."

All groups, including Group C, were then given 4 min. in which to record the experiences not asked for the first time, i.e., P if U had been asked for the first time, and vice versa. Groups A and B then ranked these in intensity.

The *E* then said to Group C only: "Now please rank *all* of these experiences in intensity from the most intense, as number one, to the least intense. For example, if the pleasant 'date' last Friday night was the most intense experience you have recalled, it would be ranked number one. If the unpleasant history bluebook was the second most intense, it would be ranked as number two, etc."

Fourteen days later, two sets of papers were prepared having the same identifying numbers as those used on the original papers. One set was blank while each paper of the second set contained all of the original responses made by a specific *S* in random order with P and U responses intermixed. Two additional experiences were added to those supplied previously by *S*. These were modified in all cases to fit the general tense and form of the other responses on each individual *S*'s paper.

The instructions to all *S*s were as follows: "Come forward and find the paper which has your number on it (offering a stack of papers that were blank except for the identifying numbers). First you will be given 2 min. in which to list all of the (pleasant or unpleasant) experiences you recorded originally that you can now recall." (Pause of 2 min.) "Now please list all of the (unpleasant or pleasant) experiences you recorded originally that you can now recall." (Pause of 2 min.)

The papers were collected and these instructions were given: "Please take the paper which

has your number on it (indicating the papers containing the listing of their original responses). The experiences which you recorded originally have been re-arranged and several others have been added. Please identify all the experiences on your list which you are reasonably certain you put on your original paper by placing a 'P' before it if you remember it *now* as pleasant or a 'U' if you remember it *now* as unpleasant. If you recognize any experiences as those you recorded but no longer feel that they are particularly pleasant or unpleasant, please mark them with an 'N.' Do not mark any experiences that you do not believe you recorded on the paper two weeks ago."

Then *E* said to Groups A and B only: "Now please rank the pleasant experiences—those you have marked 'P'—from the most pleasant as number one to the least pleasant *as you now feel about them.* (Pause.) Follow the same procedure with the unpleasant experiences, ranking them from number one as most unpleasant to least unpleasant."

The *E* said to Group C only: "Now please rank *all* experiences from most intense as number one to least intense *as you now feel about them.* For example, if the pleasant 'date' was the most intense experience, it would be ranked as number one. If the unpleasant history bluebook was the second most intense, it would be number two, etc."

Results

The utilization of personal experiences avoids the criticism that attaches to the frequent use of word lists but there still remains the question as to whether the assurance of anonymity was sufficient to permit the recall of material sufficiently related to personal life experiences to have genuine affective accompaniments. In Table II the experiences recalled are classified in an attempt to answer this question.

TABLE II

Classification of Experiences Recalled at First Recall Period

Classification	% of Experiences	% Judged Pleasant	% Judged Unpleasant
Schoolwork and grades	23.0	9.8	13.2
Family and home	6.5	5.2	1.3
Personal contacts (excluding "dates")	10.9	6.6	4.3
"Dates"	10.0	7.2	2.8
Extracurricular activities (clubs, concerts, sports)	9.6	6.8	2.8
Unclassified (abbreviated notation had meaning only to *S*)	40.0	22.0	18.0
Totals	100.0	57.6	42.4

TABLE III

Memory for Pleasant and Unpleasant Experiences within Two Contrasting Recall Sequences

Group	Pleasant Responses		Unpleasant Responses	
	Original No. (1st Recall)	% Recalled (2nd Recall)	Original No. (1st Recall)	% Recalled (2nd Recall)
Pleasant Experiences Recorded First				
A	232	40.1	158	34.8
B	403	56.1	287	48.4
C	192	56.2	122	48.4
Total	827	51.6	567	44.6
Unpleasant Experiences Recorded First				
A	148	28.4	103	44.7
B	315	42.5	279	53.0
C	245	43.3	193	61.7
Total	708	39.8	575	54.4
Combined				
Total	1535	46.2	1142	49.6

It would appear that a variety of genuine personal experiences have been tapped and that many of them have the affective accompaniments that one presumes should accompany genuine P and U experiences. The rather large number of unclassified responses resulted from our suggestion that *S* might use a word or phrase which would recall the incident to him but conceal its nature from us. It is assumed that a great many of these responses were of a personal nature, and the fact that many chose to use this protective device encourages the belief that the responses were personal and involved a substantial affective element.

TABLE IV

NUMBER OF P AND U RESPONSES ON THE SECOND RECALL

Condition	*N*	Mean	*SD*	Mean Diff.	*CR*
P responses (U requested first)	120	2.35	1.50	0.93	4.5
P responses (P requested first)	130	3.28	1.74		
U responses (U requested first)	120	2.61	1.21	0.66	4.4
U responses (P requested first)	130	1.95	1.13		

Table III gives the number of P and U experiences recalled originally and at the second recall for all groups. While analysis of such data might well avoid percentage calculations, this practice is continued because previous research with which ours is compared was reported in this manner.

The influence of temporal sequence is apparent. When P experiences were recalled first, 51.6 per cent of the previously reported P experiences were recalled. But when the U experiences were recalled first, only 39.8 per cent of the previously reported P experiences were recalled. In the case of the U experiences, although 54.4 per cent of the previously reported U experiences were recalled when U experiences were recalled first, only 44.6 per cent were recalled when U experiences were recalled second.

To test the statistical significance of temporal sequence the number of responses made by each *S* was treated as a score and the distribution of these scores for the several conditions was obtained. From these distributions, the data in Table IV were derived. The magnitude of the critical ratios indicates the importance of temporal sequence as a variable. It is hardly surprising that Jersild (5) and others who asked for pleasant experiences first should obtain results indicating the more favorable retention of pleasant experiences.

Aside from its importance in the context of experimental control, the reason for this finding can only be speculated upon. Greater recall decrement in the second half of the recall period might suggest that fluctuation of attention, fatigue, monotony, or boredom are entering in. Or it may be suggested that the first portion of the recall period sets up proactive inhibitory influences which inhibit recall during the latter portion. This experiment does not provide definitive information on this point.

Table III shows that when the effect of temporal sequence is minimized by combining all data for P experiences, whether recalled first or second, 46.2 per cent of the P experiences were recalled while 49.6 per cent of the U were recalled. It would appear that when the effect of temporal sequence is minimized, there is no clear-cut evidence that the type of affective experience is related to the ability to remember such experiences. The term "minimized" is used in recognition of the fact that the experimental groups combined to minimize

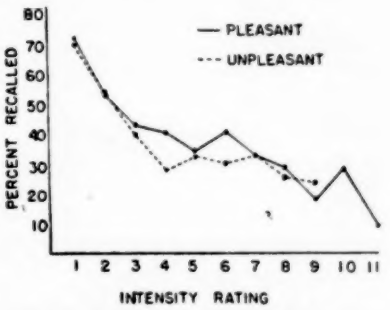

FIG. 1. Second recall as a function of intensity ratings made at time of first recall by groups A and B.

this effect were not absolutely identical in size and a negligible error results.

Figure 1 permits a comparison of the recall of P and U experiences for various degrees of intensity for Groups A and B. It is evident that the recall of both P and U experiences is markedly related to the intensity rating given these experiences by the *Ss*. Figure 1 indicates that recall drops off rapidly as one deals with experiences several intensity steps removed from the most intense experiences. The fact that the curves for P and U experiences are very nearly superimposed reflects our finding that when temporal sequence is minimized by using a counterbalanced order, P and U experiences are recalled with nearly equal facility. Figure 1 permits us to generalize this finding by showing that it holds for all levels of intensity investigated. The total number of P experiences recalled on the first recall varied from 183 for those having an intensity rank of 1 to 10 for those having an intensity rank of 11. The number of U experiences varied from 182 for an intensity rank of 1 to 4 for an intensity rank of 9.

Figure 2 depicts the relationship between incidence of recall and intensity of experience (regardless of P or U) as judged at the time of the original recall. A similar figure results when intensity, as judged at the time of second recall, is plotted against incidence of recall. The total number of experiences recalled on the first recall varied from 67 for those having an intensity rank of 1 to 11 for those having an intensity rank of 15.

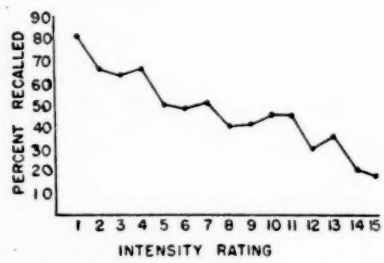

FIG. 2. Second recall as a function of intensity ratings made at time of first recall by Group C (P and U experiences rated together in intensity).

These data are of interest for several reasons. They indicate that *Ss* are capable of performing a set of operations in which they are required to disregard the qualitative nature of the experiences and attend primarily to the intensity of the experiences. Only four *Ss* had difficulty doing this, and it was due principally to a misunderstanding of directions. That primary attention was given to intensity rather than the qualitative nature of the experience was indicated by analyzing separately the intensity ratings given to P and U experiences. In this analysis, the first most intense P experience was assigned intensity rank No. 1 and the next most intense P experience was given intensity rank No. 2 regardless of the U experiences the *S* had interposed between them in making the combined ranking. The same was done for the U experiences. In this way the data were transformed so that they could be compared with the data obtained from Groups A and B where the P and U experiences had been separately ranked in intensity by the *Ss*. When this was done, the data looked approximately like those already presented for Groups A and B. This is taken as evidence that the making of intensity ranks of mixed P and U experiences, disregarding differences of qualitative nature, does no violence to judgmental processes already accepted as valid for the separate rating of P and U experiences. The effect

is one of distributing these qualitatively different experiences along the same intensity continuum without appreciably altering the intensity positions they would be assigned if they were to be judged by comparing them only with qualitatively similar experiences in the usual fashion.

This conclusion has practical implications for any attempt to determine the relationship between incidence of recall and affective tone with intensity held constant. Past attempts have characteristically shown incidence of recall for P and U experiences plotted in the same figure against a common intensity scale. This technique makes the unwarranted assumption that the same numerical values on the P and U intensity scales represent intensity equivalents. Since orthodox methodology has required the ranking of P and U experiences in separate operations, this assumption is highly questionable. The procedure described here appears to meet this objection, and Fig. 3 gives incidence of recall for P and U experiences where the intensity ratings of both types of experience were made on the same intensity scale and then intensity equivalents for P and U experiences were derived in the manner just described.

Assuming that P and U experiences alternate with each other, the mean discrepancy in intensity between similarly ranked P and U experiences plotted on the same intensity scale could be no greater than the intensity difference between adjacent ranks on the combined intensity scale. Since the procedure of ranking both P and U experiences on the same intensity scale divides the intensity range into approximately twice the usual number of divisions, the mean intensity difference between ranks on the combined scale is considerably less than

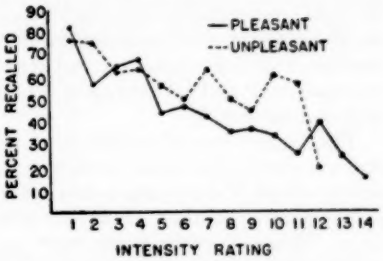

FIG. 3. Recall of P and U experiences with the factor of intensity approximately equated at each intensity level (Group C)

the difference between ranks on either the P or the U scale where intensity is ranked independently. The error is actually much less than this since many *S*s are used, and approximately half the P ranks will be half an intensity step higher than similarly ranked U experiences while half will be half an intensity step lower than similarly ranked U experiences. Occasionally, the alternation of P and U experiences on a combined intensity scale will be interrupted by a sequence of several P or U experiences. In this event, the required degree of accuracy can be achieved by utilizing only those *S*s having P and U experiences for which there is always an adjacent experience belonging to the other qualitative category. The shrinkage of data resulting from this latter step would necessitate more *S*s than were used in the present experiment.

SUMMARY AND CONCLUSIONS

Despite methodological advances in memory experiments, there remain two aspects of methodology which have not received the proper attention: (a) the importance of the temporal sequence in which P and U experiences are requested has not been adequately studied; and (b) the methods of equating the intensity of

P and U experiences have been based on the questionable assumption that similar intensity ratings on two separate scales (P and U) represent intensity equivalents.

The importance of sequence of recall was demonstrated by having some groups recall P experiences first while other groups recalled U experiences first. When P experiences were recalled first, a higher percentage of P than U experiences was recalled. When U experiences were recalled first, a higher percentage of U than P experiences was recalled. The number of responses made by each *S* under each of these experimental conditions provided score distributions which differed significantly. When the factor of temporal sequence in recall was minimized by using a counterbalanced order, there was little difference found in the ability to recall P as opposed to U experiences.

In order to equate intensity ratings of P and U experiences, it was found possible to have *S*s rate these experiences upon the same intensity scale by having them rank the most intense experience, whether P or U, as No. 1, and the next most intense experience, whether P or U, as No. 2, etc. Reasons are presented for believing that this procedure permits a much closer approximation to the establishment of intensity equivalents than does the usual procedure of rating intensity independently for P and U experiences.

Recall was found to be a function of intensity of the experience, whether P or U. With the factor of intensity approximately equated at each intensity step, the percentage of P and U experiences recalled was plotted against intensity. The resulting curves for P and U experiences differed only slightly.

(Manuscript received October 2, 1950)

References

1. Barrett, D. M. Memory in relation to hedonic tone. *Arch. Psychol.*, N. Y., 1938, **31**, No. 223.
2. Cason, H. The learning and retention of pleasant and unpleasant activities. *Arch. Psychol.*, N. Y., 1932, **21**. No. 134.
3. Flügel, J. C. A quantitative study of feeling and emotion in every day life. *Brit. J. Psychol.*, 1925, **15**, 318–355.
4. Gilbert, G. M. The age difference in the hedonistic tendency in memory. *J. exp. Psychol.*, 1937, **21**, 433–441.
5. Jersild, A. Memory for the pleasant as compared with the unpleasant. *J. exp. Psychol.*, 1931, **14**, 284–288.
6. Kowalewski, A. *Schopenhauer und seine Weltanschauung.* Berlin: Halle, 1908.
7. Menzies, R. The comparative memory values of pleasant, unpleasant and indifferent experiences. *J. exp. Psychol.*, 1935, **18**, 267–279.
8. O'Kelly, L. I., and Steckle, L. C. The forgetting of pleasant and unpleasant experiences. *Amer. J. Psychol.*, 1940, **53**, 432–434.
9. Sharp, A. A. An experimental test of Freud's doctrine of the relation of hedonic tone to memory revival. *J. exp. Psychol.*, 1938, **22**, 395–418.
10. Stagner, R. The redintegration of pleasant and unpleasant experiences. *Amer. J. Psychol.*, 1931, **43**, 463–468.
11. Thomson, R. H. An experimental study of memory as influenced by feeling tone. *J. exp. Psychol.*, 1930, **13**, 462–467.
12. Waters, R. H., and Leeper, R. The relation of affective tone to the retention of experiences of daily life. *J. exp. Psychol.*, 1936, **19**, 203–215.
13. Wohlgemuth, A. The influence of feeling on memory. *Brit. J. Psychol.*, 1922–23, **13**, 405–416.
14. Young, P. T. A study upon the recall of pleasant and unpleasant words. *Amer. J. Psychol.*, 1937, **49**, 581–596.

MEDIATED STIMULUS GENERALIZATION AS A FACTOR IN SENSORY PRE-CONDITIONING

BY DELOS D. WICKENS

Ohio State University[1]

AND

GEORGE E. BRIGGS

University of Wisconsin

The paradigm for sensory pre-conditioning experiments is as follows: The *S* is simultaneously stimulated with two neutral stimuli (e.g., light and tone). Following a number of such pairings one of these stimuli is made the conditioned stimulus for some response. Then in the crucial test the other or "non-conditioned" stimulus is presented. A majority of *Ss* so treated will give a conditioned response to the "non-conditioned" stimulus, while control *Ss* who do not receive the pretraining fail to do so (**1, 2, 3, 4**).

Taken at face value, these experiments would seem to offer evidence against an S-R approach. They appear to do so because the pre-conditioning procedure seems to consist solely of an association of stimuli. The learning would appear to be of an S-S nature rather than an S-R nature. The S-R theorist may point out, however, that some response must be made to these paired stimuli. What is even more significant is the fact that, because the stimuli occur simultaneously, the identical response or responses are given to both stimuli. The responses will vary from trial to trial, but on any one trial the response that occurs following one stimulus also occurs following the other.[2] If this be true then sensory pre-conditioning becomes a special case of mediated stimulus generalization (**6**). The transfer that occurs may be considered as arising not from the fact that the two stimuli had been associated together, but because a common and mediating response had been learned to both.

An indirect method of testing this hypothesis is to train *Ss* to make the same response to the two stimuli presented in isolation and compare the transfer value with that obtained from *Ss* who make the same response when the stimuli are presented together. The hypothesis would predict that these two groups of *Ss* would behave in essentially the same fashion in the crucial tests. The control *Ss* would be required to learn a different response to the two stimuli presented in isolation. These *Ss* should show few if any responses in the crucial tests. This is the method of evaluating the hypothesis which was employed in the present experiment. In brief, the problem was to investigate the hypothesis that sensory pre-conditioning may be a special case of mediated stimulus generalization.

METHOD

Subjects.—Forty students from the elementary psychology classes at Ohio State University

[1] This research was supported in part by a grant from The Graduate School of The Ohio State University.

[2] This theoretical analysis, although it is S-R in nature, does not necessarily assume reinforcement as essential for S-R formation. It is quite possible that the reinforcement theorist could find evidence of reinforcement in the pre-conditioning procedure. The present experiment does not deal with this problem in any way.

served in this experiment. They were selected at random from the class population and assigned at random to each of the four groups employed in the experimental design.

Apparatus.—The *S* was placed in a semi-soundproofed room which was adjacent to the control room. The experimental room contained a standard classroom chair with a finger avoidance conditioning apparatus placed on *S*'s right and a light source placed at approximate eye level when *S* was seated. A 60-w. light provided illumination during the instruction periods only.

The palm of *S*'s right hand was strapped to a large copper plate and the middle finger of that hand rested upon another small copper platform. During the conditioning trials, this arrangement formed a circuit which would result in the receipt of an electric shock if *S*'s finger did not fly up and break the circuit.

A standard telegraph key was placed in an inverted position just above the finger which was resting on the small copper plate. If *S* raised his finger approximately ½ in., the key was closed and no shock could be transmitted through the circuit. Breaking the circuit stopped a Standard Electric Time Clock, which started with the onset of any stimulation. This chronometer measured response latency in units of 1/100 sec.

The tone had a frequency of 1,000 cps and was produced by a Jackson Audio-Oscillator, Model No. 655. This tone was transmitted to *S* through the headphones. The intensity was approximately 40 db above normal threshold. The light stimulation was initiated from a DC power source which activated a 6–8 v. bulb. This power source also supplied the shock used in the conditioning trials, and this shock was variable up to 100 v. An electronic delay apparatus controlled the temporal relationship between the CS and the UCS.

General procedure.—The *S* was first brought into the experimental room, seated, and given the pre-conditioning training to the light and the tone. This constituted step one. In step two, *S* was given 30 finger avoidance-conditioning trials with the tone as CS and the shock as UCS. The third step consisted of ten trials in which the light was presented alone. This is, of course, the crucial test period for sensory preconditioning. The fourth and final step which followed immediately consisted of ten additional trials to the tone. This last step was to test for possible extinction of the conditioned response. The total time required for the entire session was approximately 45 min.

Pre-conditioning period.—The treatment of the *S*s in each group was exactly the same with the exception of step one, the pre-conditioning training. Group I made a common response to the stimuli when they appeared simultaneously in the pre-conditioning series. Group II made a common response to each of the two stimuli when they appeared separated in time. Group III made the response to only one of the stimuli (the light) and not to the other (the tone). Group IV made the response to the tone and not to the light, thus making this group comparable to the third group in that the response was made only to one of the two stimuli when they appeared separately in time.

The common response used by all groups was the verbalization of the word "now." The instructions given to Group I for the pre-conditioning period were as follows:

> This first section of the experiment is what is called a reaction time experiment. During this section, this light and a tone from the headphones which you will be wearing, will come on at exactly the same time. *Just as soon* as you are *sure* that you see the light *and* hear the tone, I want you to respond by saying the word "now" in a loud enough tone of voice so that I can hear you in the next room. This will give me an indication of your speed of reacting to *each* of these two stimuli, or in other words, I will have a measure of your reaction time. This will take approximately 5 min. and then I will give you the instructions for the second part of the experiment.

The instructions given to the remaining groups were the same except for the modification required by the difference in treatment. Group II was told that they would hear a light sometimes and a tone other times and that they were to respond to either. Groups III and IV were told to respond to one of the stimuli and not to the other.

There were 15 trials of the light and tone for Group I, and for Groups II, III and IV there were 30 trials, 15 tones and 15 lights. As indicated in the instructions, it was assumed that *S*s of Group I would make the 15 responses to *both* the light and the tone, and thus Groups II–IV were assumed to be comparable in terms of the number of reactions made to each stimulus.

The interval between trials in the pre-conditioning section varied randomly from 12 to 24 sec. In Groups II–IV the light or tone were presented randomly. No actual measure of reaction time was taken during these trials.

Conditioning period.—Following the last trial of the first session, *E* returned to the experimental room and read the following instructions to all *S*s.

> This part of the experiment is to be a conditioning study. During the experiment,

your right hand will be strapped to this electrode, and your middle finger will rest on this electrode underneath the key.

You will receive a shock which will pass from the front electrode to the one in the rear through your finger. However, you can break the shock by lifting your finger from the front electrode and closing the key in this manner.

During the experiment, a tone will precede the shock by a regular interval of time, and if you get conditioned, you will develop a tendency to respond to the tone before the shock goes on. I do not want you to fight against becoming conditioned, and at the same time, I do not want you to respond voluntarily to the tone. If your finger wants to fly up, just don't inhibit it.

The *S*'s right hand was then placed in the apparatus and the key was adjusted so that about ¼ in. upward movement closed the circuit. The *E* then left the room and the conditioning trials were begun. The time intervals between the trials for conditioning varied in a random manner from 15 to 45 sec. Thirty reinforcements were given.

Test for sensory pre-conditioning.—The final test period consisted of ten presentations of the light alone, the stimulus which had not been directly employed in the conditioning process. This was followed by ten final trials to the tone, the directly conditioned stimulus. There was no temporal break between the conditioning and the test series. The time intervals for this test were exactly the same as for the conditioning trials.

Results

In Table I is presented the mean number of responses during the crucial test to the non-conditioned stimulus, and standard deviations for each of the four groups. All differences between means are significant beyond the .01 level except for the differences between Group I and II ($t = .34$) and between Groups III and IV ($t = 2.04$).

TABLE I

Means and *SD*'s of Responses Made during the Transfer Test

Pre-conditioning Group	*N*	Mean	*SD*
I. Tone-light together	10	7.6	3.23
II. Tone-light separately	10	7.1	2.98
III. Light only	10	2.1	2.77
IV. Tone only	10	0.2	0.40

Discussion

It was predicted that the group which made the response "now" to both stimuli when they appeared simultaneously would give significantly more evidence of sensory pre-conditioning than would Groups III and IV where this response was made to but one of the two stimuli. This was confirmed at the .01 (I–III) and the .001 (I–IV) levels of confidence.

Secondly, it was predicted that sensory pre-conditioning would occur with high frequency, even though the stimuli were separated in time if the the same reaction were made to both stimuli (Group II). The results support the prediction in that (a) there is no appreciable or significant difference between Group II and Group I, and (b) in that the frequency of conditioned responses in Group II is significantly greater ($p = .01$ and .001) than in Groups III and IV where the stimuli were also separated in time but a common response was not made to both stimuli.

The procedure employed for Group II is the paradigm for mediated stimulus generalization (**5**). This involves training *S* to make the same response to two independent stimuli; conditioning another response to one of these stimuli; then testing for the occurrence of this newly trained response to the other stimulus. Our logical analysis has pointed out that in the typical sensory pre-conditioning situation these conditions are met, save that in this instance the stimuli are presented together rather than independently. This analysis, together with the empirical results, offers some justification for the interpretation of sensory pre-conditioning as a form of mediated stimulus generalization.

Summary

Four groups of *Ss* were run to test the hypothesis that sensory pre-conditioning is a special case of mediated generalization.

The procedure was as follows: For Group I a tone and light were paired together for 15 trials and *Ss* were required to say "now" when the stimuli appeared. For Group II the two stimuli were presented independently (separated in time) and the same response ("now") was given to both stimuli. Fifteen presentations of each stimuli were given. For Groups III and IV the verbal response was given to one, but not to the other stimulus.

Groups I and II showed a high frequency of response transfer or sensory pre-conditioning, and they did not differ significantly from each other. Groups III and IV showed a low level of response transfer. These groups did not differ significantly from each other, but both groups differed significantly from Groups I and II.

The empirical results support the hypothesis which views sensory preconditioning as a special case of mediated stimulus generalization.

(Manuscript received October 20, 1950)

References

1. Brogden, W. J. Sensory pre-conditioning. *J. exp. Psychol.*, 1939, **25**, 323–332.
2. Brogden, W. J. Tests of sensory pre-conditioning with human subjects. *J. exp. Psychol.*, 1942, **31**, 505–517.
3. Brogden, W. J. Sensory pre-conditioning of human subjects. *J. exp. Psychol.*, 1947, **37**, 527–539.
4. Chernikoff, R., and Brogden, W. J. The effect of instructions upon sensory preconditioning of human subjects. *J. exp. Psychol.*, 1949, **39**, 200–207.
5. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.

THE RÔLES OF PATTERN AND APPARENT DISTANCE IN DETERMINING THE COLOR OF AREAS SEEN THROUGH TRANSPARENCIES

BY W. C. H. PRENTICE, JOSEPHINE KRIMSKY, AND STEPHEN BARKER

Swarthmore College

One of the recurrent problems in the visual perception of color, has been the apparent anomaly of color "transformations" under conditions of transparency. The world seen through rose-colored glasses is said to maintain a remarkably normal look after all. And in the laboratory situation, if we examine the apparent color of a yellow disk seen through a film of blue (which may be produced either by a gelatin or by an episcotister) we see what seems to be a violation of the laws of color mixture. That is, the yellow and the blue may be chosen so as to be complementary, and yet the yellow patch still appears yellow while it is being retinally mixed with the blue of the filter or episcotister.

Nearly 30 years ago, Fuchs (**1, 2**) published two papers in which he discussed certain figural influences upon this kind of transformation and allied phenomena. Five years later, Tudor-Hart (**4**) investigated certain aspects of what Fuchs called "assimilation" with what purported to be quantitative methods. Fuchs' own investigations had been almost entirely matters of qualitative observation and description. Finally, in 1933, Heider (**3**) published a series of studies in which she reported further on the assimilation effect.

Assimilation is the tendency for a particular patch of color to take on the appearance proper to other portions of a larger figure of which the patch is a part. In the case under consideration, a yellow surface partially masked by a blue transparency, that part of the yellow that is unobscured is said to give a yellow appearance to the (retinally gray) part seen through the blue. Heider reports evidence that the assimilation effects are stronger the stronger the organization of the figure in which the patch is imbedded.

She also reports that the "yellowness" of the figure-color may be enhanced by increasing the separation between the plane of the transparency and the plane of the figure. This is a kind of "constancy" effect, said to be independent of assimilation. The fact that one portion of the visual field can be seen to lie behind another contributes, it is said, to seeing the objective (rather than the retinal) colors. The present paper reports the results of attempts to repeat Heider's experiment with more carefully defined and controlled conditions.

Reading Heider's article today, it is quite clear that the experiments are "quantitative" only in a very limited sense of the word. A portion of the article concerns experiments which involved no measurement of any sort. The cases in which some form of measuring was actually carried out are, moreover, subject to various questions and criticisms. The number of *O*s is usually too small; the *O*s are frequently not naive; and the statistics published are not sufficient to enable the reader to form any judgment about the significance of differences. Finally, the measures are all stated in terms of the percentage of blue and yellow that satisfied the *O*,

despite the fact that the particular blue and yellow used were not equal in brightness; the result being that brightness judgments and hue judgments were necessarily confounded. For these reasons, and also for certain theoretical reasons, we determined to investigate the original contentions of Fuchs, Tudor-Hart, and Heider to see whether we could reproduce the results when we paid more careful attention to such factors as stimulus control in hue and intensity, sampling of *O*s, and adequate psychophysical methods.

Procedure

The *O* sat and looked successively through two holes, each 3½ in. in diameter, in a gray screen placed some 2 ft. in front of his eyes (see Fig. 1). Behind one hole was an adjustable color wheel with the standard Maxwell disc arrangement and a protractor on the back enabling *E* to shift rapidly and accurately from one setting to another. Behind the other hole in the screen stood a blue episcotister which could be raised or lowered. Through the episcotister a yellow figure (equal in brightness to the blue of the episcotister) could be seen 4 ft. away on a gray background.

Four different arrangements were used. In Condition A, yellow arcs were pasted on the episcotister in such a way as to form part of a circle whose center was the center of the episcotister.

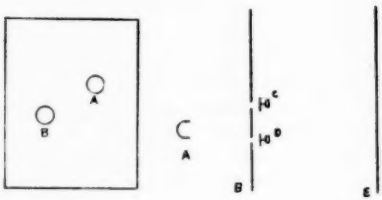

Fig. 1. Part 1a is the screen that confronted *O:* he saw at A the various combinations of yellow figure and blue episcotister and at B a Maxwell disc color mixer. Part 1b is a schematic view of the experimental room: *O*'s chair is at A; the screen of Part 1a is at B; C and D represent the color mixer and the episcotister, respectively; and *E* represents the screen on which the figures for observation were hung.

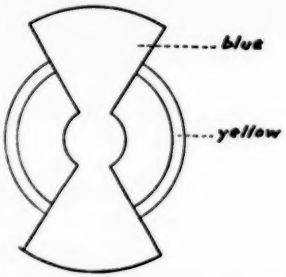

Fig. 2. The Wallach episcotister. Arcs of cardboard are pasted into place as shown so that the entire assembly can be spun on a color wheel, giving the same retinal stimulation as is given by a full circle of yellow seen through a blue episcotister.

This arrangement (which we will refer to hereafter as the Wallach episcotister) is shown in Fig. 2. The gray screen, 4 ft. away, remained blank in this condition. In Condition B, a yellow ring was presented on the distant screen completely covered by the area of the episcotister, so that *O* saw an annulus filled with and surrounded by blue in this condition as in Condition A, the only difference being that in Condition A, the annulus and surround lay in the same plane, and in Condition B, the annulus was visibly lying on a surface 4 ft. behind a blue transparency. In Conditions C and D, the episcotister was lowered so that it filled only half the opening of the near screen. In C, the ring of Condition B was again used, half of it being seen through the blue episcotister and half, unobscured, above. In D, two yellow arcs, both curving the same way, were used, one appearing through and the other above the episcotister. In both C and D, it was apparent that the yellow figures lay in the same plane.

The color wheel was composed of the same yellow and blue papers of equal brightness (Munsell 5/5-Y and 5/5-PB). The episcotister was set at 252° blue, so that when the yellow figure behind the blue was viewed through a reduction screen, it looked like a slightly bluish gray. It was intentionally made slightly bluish in order to counteract some of the contrast effect, since a gray figure surrounded by blue will always take on some degree of yellow appearance. This tendency, operating as it does to give results like those predicted on grounds of transparency effects or assimilation, must be given special consideration and control. The daylight illumination on each of the experimental surfaces (the screen in front of *O*, the Maxwell disc, and

the screen at a distance of 4 ft.) was equated by means of a Macbeth illuminometer.

Conditions A and B were introduced to test the effect on color of transparency, i.e., of having the same area on the retina represent two surfaces of different colors and lying in separate planes. In A there could be no separation of planes, but the same color mixture stimulated the retina in this and the other conditions. Because of the 4-ft. distance between the episcotister and the screen, a very good transparency condition existed in Condition B, and *O* saw the yellow ring lying four feet behind the plane of the blue transparency. Conditions C and D served to test the effect (on apparent hue) of two different degrees of figural cohesiveness between inlying and outlying portions of the figure. According to Heider's analysis of the situation, the ring should be yellower in Condition B than in Condition A; and in C the portion seen through the episcotister should be yellower than the less cohesive figure of Cond. D. Of all the conditions, Cond. A should presumably be the bluest.

The *O* looked at the opening in the near screen and saw a yellow circle surrounded by blue (in Cond. C and D a yellow *semi*-circle surrounded by blue and another outside the blue, surrounded by gray) and then had to judge whether the yellow was more or less yellow than a comparison mixture seen through another hole in the same screen.

Fifteen *O*s were used. The order of Cond. A, B, C, and D was counterbalanced for different *O*s in order to control possible adaptation effects due to the length of time of the observations. Each *O* made a total of about 30 judgments for each condition in the course of determining six limits (using 10° step-intervals), three from blue to yellow and three from yellow to blue in ABBAAB order. A two-category response ("too blue" or "too yellow") was required at all times. Between judgments, *O* looked away from the screen in order to minimize adaptation and satiation effects and to prevent him from seeing the new setting on the color wheel before it was started spinning. Most *O*s had had some previous experience making psychophysical judgments. Most, but not all, *O*s knew something about the purpose of the experiment; no differences are apparent in the performance of our sophisticated and our few naive *O*s.

Results

In Table I are presented the mean number of degrees of blue (six limits for 15 *O*s) required to match the hue of the "yellow" figure under each condition. None of the differences approaches statistical significance. If one anomalous *O* is omitted, the results become even more nearly the same, namely 132.6, 131.9, 132.7, and 132.3 degrees under Cond. A through D, respectively.

TABLE I

Mean Number of Degrees of Blue Required to Match the Hue of the "Yellow" Figure under Four Conditions

Condition	Mean	*SD*
A. Wallach episcotister	128.7	20.5
B. Obscured ring	133.8	20.0
C. Half-obscured ring	132.7	16.9
D. Broken ring	132.3	19.0

Discussion

According to these results, it would appear that under our conditions no color constancy occurs beyond the yellowing predicted from knowledge of the contrast conditions in the control Cond. A. It may be that this difference between our findings and those of Fuchs and Heider should be attributed to the larger number of *O*s used here, as well as to more careful control of illumination and brightness. Preliminary experiments have shown that any difference in the last variables may cause a significant difference in color judgments. It can be seen that in Cond. A, where there is no possible separation of planes, the color wheel setting is about 130° of blue, while the yellow figure with which it is being equated is actually seen through episcotister segments totaling 252°. This difference of over 100° appears to result from the well-known influence of simultaneous contrast between the "yellow" ring and the inducing blue field of the episcotister. In Cond. B the same difference exists, and without Cond. A as a control

might have been explained by Fuchs and Heider as an example of color constancy or transformation due to the separation of planes. The usual method of getting rid of the separation of planes has been to use a reduction screen. Actually, however, what this accomplishes is more than just the pulling together of the colors into one plane. It also succeeds in excluding the blue surround of the mixture color. In our Cond. A, we have kept everything constant except the one factor of the separation of planes. It is now perfectly clear, as indeed an experiment of Wallach (5) indicated earlier, that the condition of transparency does not affect the apparent hue of the figure, which is instead wholly determined by contrast influences.

The question remains, how did it happen that Heider thought she had found a result that cannot be repeated under our conditions? The most reasonable explanation appears to be that she probably did not maintain precisely equal illuminations on the two surfaces. That is to say, the match which an *O* will make when looking at our Cond. B will be the same as that in Cond. A *only if the total illumination on the two yellow rings is the same.* The apparent hue of the yellow ring is especially sensitive to very small changes in illumination. We have found, for example, that rather minor adjustments in the position of a window shade will make the yellow ring appear either much yellower or much grayer. If, for some reason, Heider had a set-up in which the farther surface was more highly illuminated, the proper proportional intensities of blue and yellow on the retina would obviously have been disturbed.

It is rather more difficult to explain how Heider was able to get her reported results concerning the relationship between the cohesiveness of the figure and the apparent hue. Her own data are not particularly convincing in this respect. She failed to report the data in a way which would permit estimation of the significance of differences, and it is certainly not apparent that the differences she reports are statistically significant. Since we have been unable to find even the small differences that she reports, we are inclined to believe that Heider's original results are a matter of sampling error of some sort.

One possible explanation of the discrepancy should be explored. In the well-known Benussi circle (Fig. 3), assimilation (i.e., uniformity of color in the ring) is achieved when the ring is unbroken by any sharp contour. It is possible that Heider's conditions and ours may have differed in this respect. While it is true that the edge of the episcotister would ordinarily provide such a contour and clearly did so in our conditions, there is one situation in which that might not be the case. If the brightness of Heider's background (against which the yellow figures appeared) was identical to the brightness of the episcotister surface, then contour formation would have to be achieved by hue differences alone. There is no doubt that this would blur the contour to some extent. Whether that fact alone would suffice to account for the difference in our two results is a matter for investigation.

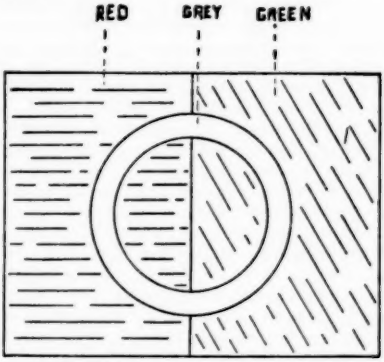

FIG. 3. The Benussi circle. When a heavy contour breaks the gray circle along the line that separates the two background colors, the simultaneous contrast effect is enhanced dramatically so that the pinkish tinge of the right-hand half of the ring and the greenish tinge of the left become very strong instead of barely noticeable.

Another possibility is that *O*s who are naive or *O*s who have not had a good deal of experience making psychophysical judgments, might easily be persuaded that the colors under the two conditions (solid circle and broken circle) are different. It is not easy to judge the portion of the total distant field seen through the episcotister and judge it independently of that portion of the same yellow paper that is seen outside the field of the episcotister. When one is confronted with the uncontaminated yellow and the yellow-blue mixture simultaneously, it is very easy to "interpret" the yellow-blue mixture as being a yellow paper seen through a blue screen. Under these conditions, if the matching experiment is not carried out carefully, *O* may very easily get the impression that this judgmental reaction is a perceptual one. It is only when he compares the part of the yellow field that is seen through the episcotister directly to the Maxwell disc that it can be shown that what he sees is not different from what he sees in the case where the entire yellow field resides behind the episcotister. A naive *O* who is first shown the yellow ring seen entirely through the episcotister and who is then shown the yellow ring slid up until half of it appears beyond the edge of the episcotister is likely to react by saying "Oh." This reaction is a matter of his sudden recognition that the yellow ring is really a much more saturated yellow than he had realized before, although he had recognized it as yellow rather than gray from the beginning. Under the influence of this attitude, the *O* may change his interpretation of the gray portion of the total field and attend to it as a portion of the now yellower circle merely obscured by a blue mist or film. We considered the possibility, in a sense that is rather difficult to define or specify, that the bottom part of the circle might now somehow *look* yellower for some *O*s than it did before. This possibility would be even more likely if *O* had some expectations or theoretical presuppositions about the conditions beforehand.

The procedure of our first experiment might actually obscure such a result by requiring *O* to make a long series of somewhat artificial judgments. As a consequence, we set up a simpler test. In the second experiment, the same conditions of observation were used, but a gray field containing two small (2-in. diameter) yellow discs was exposed. One of the discs was seen only through the episcotister; the other was partially obscured by the episcotister and partially exposed to direct view. The *O*s were asked to make a simultaneous comparison between the two yellows seen through the blue film. Six naive *O*s reported that they could see no difference. Three sophisticated *O*s gave exactly the same result.

Summary

Using careful quantitative methods by which *O*s judged the apparent yellowness of a figure seen either through or surrounded by a blue of the complementary hue, we have been unable to find any influence of the separation of planes or of assimilation. Earlier authors had reported a distinct enhancement of the contrast effect under conditions of transparency and of "strong figures." It is suggested that their results were due to inadequate control of illumination or sampling error.

(Manuscript received October 21, 1950)

References

1. Fuchs, W. Experimentelle Untersuchungen über die Änderung von Farben unter dem Einfluss von Gestalten ("Angleichungserscheinungen"). *Z. Psychol.*, 1923, **91**, 145–235.
2. Fuchs, W. Untersuchungen über die psychologischen Grundprobleme der Tiefenwahrnehmung. *Z. Psychol.*, 1923, **92**, 294–325.
3. Heider, Grace, M. New studies in transparency, form, and color. *Psychol. Forsch.*, 1933, **17**, 13–55.
4. Tudor-Hart, Beatrix. Studies in transparency, form, and colour. *Psychol. Forsch.*, 1923, **10**, 255–273.
5. Wallach, H., and Galloway, A. Constancy of colored objects in colored illumination. *J. exp. Psychol.*, 1946, **36**, 119–126.

THE ABILITY TO MAKE ABSOLUTE JUDGMENTS OF PITCH [1]

BY ALICE VAN KREVELEN [2]

University of Rochester

The ability known as absolute pitch has been a subject of interest for many years. Neu (6) recently reviewed the work of 58 authors of which, however, comparatively few used *Ss* already possessing the requisite absolute ability. Petran (7) in a study resembling the second experiment in the present investigation did observe nine *Ss* with absolute pitch. His article reviews the literature up to 1932. Bachem (1, 2, 3, 4) has made the most recent investigation of known possessors of absolute pitch. Nevertheless, a more precise, quantitative report of the performance of people possessing this ability seems necessary in view of the meagre data describing their accuracy in any finer units than semitones.

Problem

This investigation was an attempt to determine how accurately judgments of pitch could be made by persons definitely possessing the ability known as pitch recognition or absolute pitch. An unusual degree of precision in frequency monitoring was utilized in an attempt to discover the limits of the ability.

Two experiments were performed. In one experiment *Ss* were required to identify randomly presented oscillator tones which covered the range from 404 to 478 cps in 2-cps steps. This will be referred to as the "recognition method." In the second experiment the same *Ss* attempted to adjust oscillator frequency so as to produce *G*♯ (415.30 cps), *A* (440 cps) and *A*♯ (466.16 cps). This is called the "production method."

Previous studies have shown that musical *Ss* judge best those tones lying in the middle range of the musical frequency scale. Since it was desired to get a measure of the very best performance that could be expected, the three tones to be investigated were selected from this middle range. It is unfortunate that only a very small number of frequencies could be thoroughly studied within the amount of time which the *Ss* could reasonably be asked to give.

Subjects

The *Ss* were chosen according to Bachem's criterion for universal genuine absolute pitch (1). All students in the Eastman School of Music during the academic year 1948–49 who had given evidence of absolute pitch ability by identifying piano tones on an entrance test were asked to report for a second test. For this group test all *Ss* were required to identify all 88 tones on the piano. Each key was held down for approximately 1 sec. and the time between notes was just sufficient for *Ss* to write down the proper tone if recognized immediately. [3]The papers were collected as soon as the last note was given so that there was no time for *Ss* to check over their answers. The sequence of tones was arranged to rule out the effects of relative pitch as far as possible: no two consecutive tones were taken from the same octave and the smallest

[1] This paper is a portion of a dissertation submitted to the Department of Psychology of the University of Rochester in partial fulfillment of the requirements for the Ph.D. degree. The writer is indebted to Professor Edward H. Kemp of the Department of Psychology and to Professor Wayne Barlow of the Eastman School of Music for advice and assistance throughout the investigation. The research was supported in part by funds from the American Otological Society. The Conn chromatic stroboscope used in this investigation is the property of the U. S. Navy, and was loaned for the experiment.

[2] Now at Hollins College, Virginia.

intervals were those usually considered most difficult to identify, such as major and minor ninths, augmented elevenths, fourteenths, etc. Twenty-three *Ss* were selected from the 40 who took the test. These were *Ss* who either identified all 88 notes correctly or who made occasional half-tone errors in the extreme upper and lower octaves, thus conforming to Bachem's criterion.

This group was given a third test with pure tones produced by means of an oscillator and headphones. This was an individual test in which *S* gave his judgments aloud. These were recorded by *E*. Since the experimental tones were not to be taken from the extremes of the scale only the 48 tones lying in the middle four octaves were used. The sequence of tones was again arranged to rule out the effects of relative pitch as far as possible.

The 17 *Ss* (10 women and 7 men) who identified all 48 oscillator tones correctly were selected for the study. The age range of these *Ss* was from 17 to 37 years with the majority between the ages of 17 and 22. All *Ss* filled out a questionnaire before the experimental sessions began. They gave information concerning their musical history, family musical background, etc. (Table I). The mean age of beginning formal music study was five years and all the *Ss* were engaged in the professional study of music with the intention of pursuing it as a career.

TABLE I

DESCRIPTION OF SUBJECTS

S	Age	Music Degree Sought	Major Instrument	Age Began Music Study	Musical Parents*
1	21	B.M.	Harp	5	2
2	17	B.M.	Piano	4	1
3	18	B.M.	Piano	7	1
4	37	Ph.D.	Organ	6	1
5	26	M.M.	Piano	5	2
6	18	B.M.	Piano	6	2
7	21	M.M.	Voice	3	2
8	21	B.M.	Piano	5	2**
9	21	M.M.	Piano	4	1
10	22	B.M.	Piano	5	1
11	19	B.M.	Piano	4	0
12	18	B.M.	Piano	4	1
13	19	B.M.	Piano	5	1
14	17	B.M.	Horn	2½	1***
15	20	B.M.	Piano and voice	2½	1
16	33	Ph.D.	Clarinet	10	0
17	22	M.M.	Voice	5	0

* Includes both amateur and professional.
** Also had an aunt with claimed absolute pitch.
*** Also had a brother with claimed absolute pitch.

APPARATUS AND PROCEDURE

Experiment I.—The experiment was carried out in a small room having sound-deadening wall panels and protected from hall noises by two closed doors while the sessions were in progress. The room was in an unusually quiet location, and at no time were there any musical sounds audible during an experimental session.

The *S*, wearing Permoflux PDR-8 earphones mounted in sponge rubber, sat at a small table shielded from the *E* and equipment. The tones used were produced by a Hewlitt-Packard Audio-Oscillator, Model 200D. Sensation threshold was determined for each *S* at 422 cps, which was midway in the range of frequencies to be used, and all tones were presented at an intensity 40 db above this level. The output of the oscillator was connected through an attenuator into the earphones while a Conn chromatic stroboscope was used to monitor the input to the attenuator (8).

The *E* tuned the oscillator while watching the stroboscope pattern until the desired frequency was being produced. She then pressed a telegraph key which led the tone to the earphones of *S* and held the key down for approximately 1 sec. No attempt was made to eliminate a click because the duration seemed long enough to minimize the disturbing effect of transients. Approximately 10 to 15 sec. elapsed between tones in the series.

Tones were presented to *S* in random order and the same frequency appeared ten times in the series. The tones range from 404 cps, about a quarter tone below $G\sharp$ to 478 cps, a quarter tone above $A\sharp$, in 2-cps steps. Two exceptions had to be made, however, in order to include the correct frequency for both $G\sharp$ and *A*, one being an odd number of cycles and the other an even. The total series consisted of 390 tones. The order was prearranged according to a table of random numbers and was the same for all *Ss*. The *S* received 100 tones at each session.

The following instructions were given at the first session and *S* was briefly reminded of them at the beginning of subsequent sessions: "You will hear a series of tones. Some are on pitch and some are not. We are interested in finding out just what you will accept as sounding like a good *C*, $C\sharp$, etc. For instance, suppose you hear a tone lying between *D* and $D\sharp$; if it sounds closer to *D* report it as a sharp *D*; if it sounds closer to $D\sharp$ report it as a flat $D\sharp$. In other words always report in terms of the closest musical tone. If you say just *D*, $D\sharp$, etc., we will understand that the tone sounds exactly on pitch to you. Report your judgment aloud after each tone and I will write it down over here."

The *S*s were not aware of the limits of the frequency range to be investigated as it was not desired to give them any frame of reference other than a purely subjective one on which to base their judgments.

Experiment II.—The apparatus used was the same as in Experiment I. This time *S* sat at a small table next to *E*'s table but separated from it by a partition. Through an opening in the partition he could insert his hand and arm and turn the dial of the oscillator without being able to see the dial or the oscillator. The oscillator was monitored by the stroboscope as in Experiment I and the tones were presented at the same intensity. The tones requested were *G*♯, *A* and *A*♯. In the total series *S* "found" each tone 100 times, 50 times from original settings below the goal tone and 50 times from above. Five preliminary settings were chosen in 10-cps steps around a musical fifth above and below the tone being requested. Therefore, the original settings were never exactly on pitch for any tone. Each setting was used ten times in the series. The order in which the tones were requested and the original settings were randomized and the same random order was used for all *S*s.

These *S*s were given the following instructions: "This time you will be able to change the tone yourself by turning the oscillator dial. You can vary the frequency until it sounds like the tone I am asking for. One way to do this is to go to the tone you want and then make it either a little sharp or a little flat, depending on whether you are coming from above or below the tone, and then go back to where you hear the tone you want again. If you prefer to go slowly until you hear the tone you want and then stop, that is all right too. We want you to do it the way you get the best results. However, do not go as much as a half tone above or below the tone you are looking for when making final adjustments. You can have a few trials to see which way to turn the oscillator dial to make the tone go up or down and how much you have to turn to make a certain amount of change. Then I will ask you to produce a note. As soon as you hear a tone in your ear start turning the dial to get what you want. All the notes I ask for are to be in the middle-*C* octave, the same octave as the tones you listened to in the other part of the experiment. As soon as you have the tone you want remove your hand from the dial and I will know that you have finished tuning."

As soon as *E* had tuned to the desired preliminary frequency the tone was keyed into the earphones and the key was held down until *S* removed his hand from the dial to signal that he had finished tuning. The tone was then keyed out of the earphones and *E* turned the stroboscope dial until some stroboscopic frequency coincided with the frequency being produced by the oscillator. The reading was recorded in cents (.01 semitone) deviation from the goal tone, both magnitude and direction of error being noted. About 50 judgments were made at each session.

Results

Exp. I: Recognition method.—The individual means and standard deviations for all *S*s for the recognition method are presented in Table II. The means are the averages of all the frequencies that were judged by the *S*s to sound exactly on pitch for each note. It should be mentioned that 510 of the total of 6,630 tones heard by the 17 *S*s were actually on pitch for *G*♯, *A* or *A*♯. A little better than one-third (2,313) of the judgments made were "on pitch" judgments; more than four times the number of tones that were on pitch were called "on pitch." Furthermore, of the 170 true *G*♯s presented, 44 per cent were named correctly while the remainder were judged to be either sharp or flat; 47 per cent of the 170 *A*'s and 52 per cent of the 170 *A*♯'s were recognized correctly. The *SD*, in cps, was calculated around the *S*'s own mean and is an index of the consistency of his judgments. The constant error may be seen by comparing the mean with the accepted physical standard for a particular tone, based on an *A* of 440 cps.

The constant errors were larger for *A*♯ than for *G*♯ and *A* for the group as a whole. The *S*s were also less consistent in their judgments for *A*♯ than for *A* as measured by the *SD*'s. The mean *SD*'s for the group, obtained by summing the individual *SD*'s and dividing by the number of *S*s, were 7.57 cps for *G*♯, 6.40 cps for *A* and 7.11 cps for *A*♯. These *SD*'s are of the magnitude of about .34 semitone.

TABLE II

PERFORMANCE OF SUBJECTS USING THE RECOGNITION METHOD

S	Reported as *G#*		Reported as *A*		Reported as *A#*	
	Mean in cps	*SD* in cps	Mean in cps	*SD* in cps	Mean in cps	*SD* in cps
1	414.86	4.56	437.18	5.39	456.58	4.67
2	413.40	4.29	434.36	4.51	460.35	7.66
3	441.50	14.20	453.92	12.23	467.25	7.15
4	417.60	6.86	438.09	6.22	452.57	10.01
5	420.20	8.98	437.57	6.35	460.00	8.18
6	416.96	7.23	433.47	6.15	457.06	6.32
7	431.81	10.51	449.37	6.93	463.16	7.57
8	426.50	9.26	445.47	7.70	465.30	7.95
9	423.27	9.39	445.95	8.07	469.18	10.72
10	419.30	7.58	440.53	3.51	461.89	5.96
11	423.18	7.02	444.77	5.10	470.07	5.88
12	421.05	7.02	441.41	5.40	463.89	5.52
13	413.96	4.63	441.38	7.46	461.75	4.94
14	416.12	4.52	436.64	4.14	457.80	6.94
15	422.27	8.08	439.00	4.97	462.68	6.24
16	424.36	10.10	441.61	9.99	462.84	8.99
17	413.24	4.42	437.74	4.70	463.57	6.18
Mean of pooled judgments	421.08		440.43		460.30	
Standard frequency	415.30		440.00		466.16	

Figure 1 shows the pooled "on pitch" judgments for all *S*s. It will be seen that the peaks for all tones occur at frequencies coinciding with the true standards but that there is also overlap between categories and the range for any one category is quite wide if we consider all the judgments that were made. There is even a small range from about 440 cps to about 452 cps where all three categories were used in identifying the tones.

The curves, as viewed roughly, seem to represent fairly normal distributions of judgments around peaks coinciding with the true values of the particular notes. The curve for *G#* appears to be cut off at the lower end and that for *A#* at the upper end. This is probably due to the experimental conditions since the frequency of 404 cps was the lowest that was given and that of 478 cps was the highest that was given. If the frequency range had been extended in either direction it is probable that a few *S*s would still have occasionally used the categories of *G#* and *A#* and these curves would correspond with that for *A*.

Means were calculated for each category of the pooled judgments of all 17 *S*s. These means were found by averaging all the frequencies that were judged to sound exactly on pitch for each note.

Exp. II: Production method.—Table III contains the individual means and *SD*'s for all *S*s for the production method. The individual means are the averages of all the settings that were made by each *S* for a particular tone and the *SD*'s are calculated around these means. The constant errors can be calculated from the accepted physical standard for each note, based on an *A* of 440 cps.

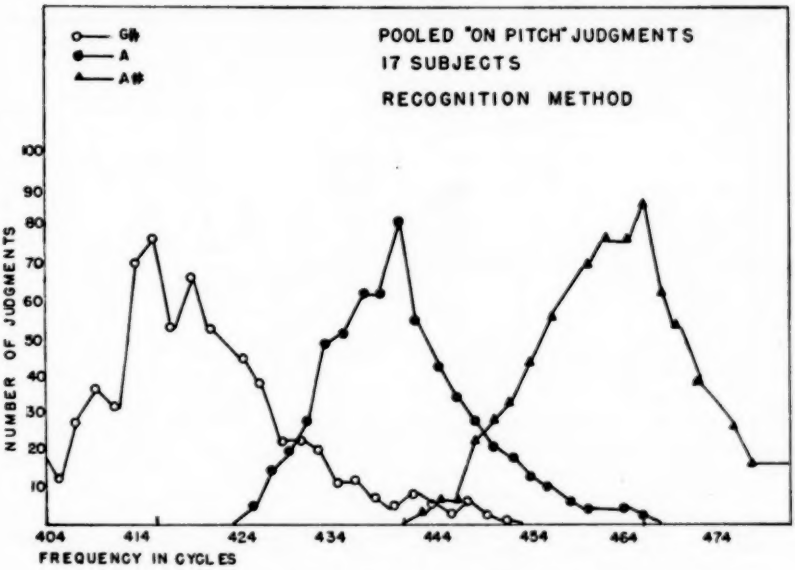

FIG. 1. Pooled distributions of actual frequencies of tones judged by 17 *Ss* to be *G♯*, *A*, and *A♯* (Recognition Method)

TABLE III

PERFORMANCE OF SUBJECTS USING THE PRODUCTION METHOD

S	Selected as *G♯*		Selected as *A*		Selected as *A♯*	
	Mean in cps	*SD* in cps	Mean in cps	*SD* in cps	Mean in cps	*SD* in cps
1	411.25	4.60	431.69	6.14	458.42	6.35
2	411.96	4.79	435.20	5.49	456.83	9.56
3	424.03	11.25	451.33	8.12	467.51	5.72
4	412.44	6.11	439.24	5.59	455.25	5.77
5	416.75	4.08	439.75	5.36	462.14	4.89
6	411.72	4.24	431.44	4.10	455.52	4.42
7	425.50	6.43	442.80	4.17	463.75	4.87
8	412.44	4.48	437.21	4.90	461.34	5.35
9	416.99	4.66	437.97	4.12	459.48	6.86
10	417.95	4.54	437.72	4.82	460.28	7.20
11	414.59	5.46	441.53	5.17	465.36	4.91
12	412.44	5.47	435.45	6.27	457.89	6.27
13	416.75	3.19	443.32	3.66	468.05	8.23
14	418.43	3.34	437.21	3.56	460.81	4.26
15	416.99	5.46	437.47	5.89	462.68	7.39
16	412.67	5.05	442.55	6.21	468.32	8.59
17	417.23	2.92	439.75	3.38	468.86	3.98
Mean of pooled judgments	415.67		438.73		460.21	
Standard frequency	415.30		440.00		466.16	

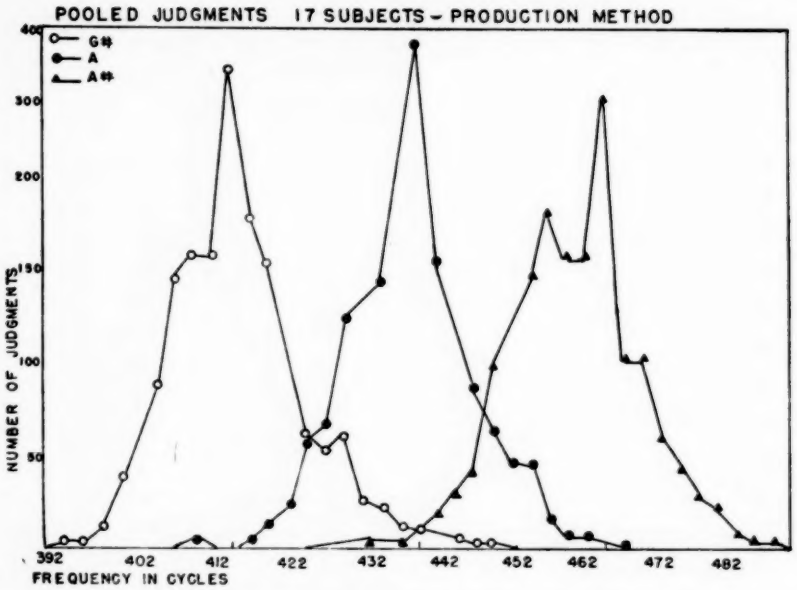

FIG. 2. Pooled distributions of actual frequencies of tones judged by 17 *Ss* to be $G\sharp$, *A*, and $A\sharp$ (Production Method)

Figure 2 presents graphically the pooled judgments of the group for the production method. Again, as in the recognition method, the judgments for each tone are distributed around a peak which coincides with the true physical standard for that particular category.

Means were calculated by pooling the judgments of all 17 *Ss*. These values differ only slightly from the values obtained when the 17 individual means were averaged.

Comparison of methods.—It may be seen by referring to Tables II and III that the individual *Ss* seem to be more consistent in their judgments (they show smaller *SD*'s) if the production method is used rather than the recognition method. The averaged *SD*'s

TABLE IV

DIFFERENCES BETWEEN MEAN *SD*'S OF JUDGMENTS UNDER THE CONDITIONS OF RECOGNITION AND PRODUCTION

Tone	Recognition Method		Production Method		Difference		
	Mean *SD* in cps	*SD*	Mean *SD* in cps	*SD*	Mean in cps	$\sigma_{Diff.}$	*t*
$G\sharp$	7.57	2.56	5.06	1.85	2.51	.429	5.85**
A	6.40	2.15	5.12	1.19	1.28	.445	2.87*
$A\sharp$	7.11	1.64	6.15	1.57	.96	.547	1.75

* Significant at the .02 level of confidence ($df = 16$).
** Significant beyond the .01 level of confidence ($df = 16$).

TABLE V

DIFFERENCES BETWEEN MEAN AVERAGE ERRORS OF JUDGMENTS UNDER THE CONDITIONS OF RECOGNITION AND PRODUCTION

Tone	Recognition Method		Production Method		Difference		
	Mean in Semitones	*SD*	Mean in Semitones	*SD*	Mean in Semitones	σ_M	*t*
G♯	.342	.231	.293	.100	.049	.044	1.11
A	.234	.115	.290	.084	−.056	.022	2.54*
A♯	.269	.090	.325	.090	−.056	.024	2.33*

* Significant at the .05 level of confidence ($df = 16$).

for the group are also smaller for the production method.

The *t*-test was applied to differences in performance under the two conditions as measured by the *SD*'s (Table IV). The formula used to find the standard error of the mean difference is that used for matched pairs (5). For *G*♯ a *t* of 5.85 was obtained. This is significant at better than the .01 level of confidence. The *t*-value for *A* was 2.87 and that for *A*♯ was 1.75, the value for *A* being significant at better than the .05 level, and the value for *A*♯ is not significant. So it can be said that for the tones *G*♯ and *A* the *S*s were more consistent in their judgments when the production method was used.

The *t*-test was also applied to differences in performance under the two conditions as measured by the average error taken from the true standard for each particular tone (Table V). The *SD* around the *S*'s own mean is an indication of the consistency of his judgments but the average error taken from the true standard is a reflection of the accuracy of his judgments. For *G*♯ the *t* was 1.11 indicating that there was no significant difference between accuracy of performance in the two experiments. The *t* for *A* was 2.54 showing a difference significant at better than the .05 level of confidence in favor of the recognition method. Likewise, the *t* for *A*♯ (2.33) was significant at better than the .05 level and also in favor of the recognition method. We must conclude that, while the *S*s were more consistent in their judgments when they could control the frequency themselves, they were more accurate in their judgments for *A* and *A*♯ when they identified the heard tones.

DISCUSSION

One of the objects of this investigation was to discover how wide a band of frequencies would be accepted as sounding on pitch for a given tone. The results which answer this question are shown graphically in Fig. 1 and 2, and have been discussed above. We conclude that no sharp judgmental limits for all *S*s for any one tone were demonstrated by this experiment.[3]

Means were calculated to give an idea of the central tendency in making judgments and for comparison with the true physical standard for each particular tone. Tables II and III give the means of the pooled judgments for all 17 *S*s for both methods. The mean for *G*♯ was higher than the standard by about 6 cps when the

[3] For a more complete analysis and description of the performance of each individual in this experiment the reader is referred to a thesis on file at the University of Rochester under the same title as this article.

recognition method was used and deviated less than 1 cps from the standard when the production method was employed. For judgments of *A* the mean values in both experiments were within approximately 1 cps of the true value. The means for *A*♯ were approximately 6 cps lower than the standard under both the recognition and production conditions. It should be noted, however, by referring to Fig. 1 and 2, that the modal value of the pooled judgments actually coincides with the true physical standards. When the 17 individual means were averaged for each method they were very similar to the means of the pooled judgments, which perhaps indicates that the group is fairly homogeneous.

Consideration of the means alone tells us nothing of the variability. The consistency and accuracy with which judgments were made is best indicated by the *SD*'s and the average errors. The mean average errors, in general, for both methods and for all tones were of the approximate magnitude of .3 semitone. Bachem (**1, 2, 4**) has reported average errors of .1 semitone for his *S*s. In the present experiment no single *S* made an average error as small as .1 semitone. One of the 17 *S*s had average errors only slightly higher than .1 semitone for three of the tones, one with the production method and the other two with the recognition method. However, the performance of *S*s in the present experiment cannot readily be compared with that of Bachem's *S*s since it is not easy to deduce from Bachem's publications exactly how he arrived at the figure of .1 semitone.

The differences, pointed out earlier, between the two methods hardly seem reasonable on first consideration. Why should the "production" method give the more consistent judgments while the "recognition" method gives the more accurate ones? One possible explanation for this result may be as follows. In the production method the *S* was never given any frequencies that were exactly on pitch for preliminary settings but arrived at his idea of the goal tone only by tuning through a series of intermediate frequencies until he reached a point which satisfied his criterion. He may have been satisfied with a setting which was some distance away from the true standard on the first setting and the very consistency with which he turned to this same tone increased his average error. When *SD*'s are compared for the two methods they are significantly smaller for the production method for the tones *G*♯ and *A*. It would seem that *S*, being able to control the frequency himself, could establish his own criterion for a tone and turn to it quite consistently.

It may be argued that only the first judgment made in any session is without reference to any other heard tone. In the production experiment even the first judgment would not satisfy this requirement since *S* tuned through several intermediate frequencies to reach his goal tone. But when *S* is given no outside cue as to the correctness of his judgments, the fact that he has just heard another "unknown" tone would seem to be of little help. Furthermore, even if it were desirable to use only the first judgment each day, the time required to obtain statistically reliable results would make such experimentation a practical impossibility.

Petran (**7**) characterized his *S*s with absolute pitch as having the ability to make the majority of their judgments using the production method, within *G*♯ to *A*♯ when they were trying to produce *A*. The 17 *S*s in the present experiment were apparently more accurate, since only occasional settings varied as much as a half-tone

from the goal tone and average errors were of the magnitude of .3 semitone. It is possible that if average errors were calculated for all settings made by Petran's best *Ss* they would have been of the same order of magnitude. Petran does not describe his method of selecting *Ss* so there is a possibility that not all of his nine *Ss* would have satisfied Bachem's criterion for genuine absolute pitch.

Bachem mentions two factors as being of importance in the development of absolute pitch: (a) an attention factor, and (b) an inheritance factor. In this group of *Ss* possessing genuine absolute pitch only two *Ss* reported any relatives with absolute pitch. Since the mean age of beginning formal music study was five years and since most *Ss* reported either one or both parents as being engaged in music either professionally or on an amateur basis, it is safe to assume that these *Ss* were hearing musical tones at a very early age. In these particular *Ss*, therefore, the factor of primary importance may have been attention to tones at an early age.

Summary and Conclusions

Seventeen *Ss*, selected according to Bachem's criterion for genuine absolute pitch, performed in two experiments. In one they were required to identify randomly presented oscillator tones which covered the range between 404 and 478 cps in 2-cps steps. In the second experiment the same *Ss* attempted to adjust oscillator frequency so as to produce *G*♯ (415.30 cps), *A* (440 cps), and *A*♯ (466.16 cps).

The *Ss* were more consistent in their judgments, as measured by *SD*'s, for two of the tones when the production method was used than when the recognition method was used. However, performance was more accurate, as measured by the average error, for two of the tones with the recognition method.

The range of individual means shows that, as judged by the performance of this group, there is no particular frequency which is agreed upon by all individuals to correspond to a particular tone; but when all judgments for all *Ss* are pooled the greatest number of judgments coincide with the true physical standard for a given tone. This would seem to indicate that individual experience is important in establishing the tonal norm for an individual and that the norm for these selected *Ss* usually falls rather near the true physical standard. Even with the established individual norms the *Ss* differed in the consistency with which they could retain their own standards through a long series of judgments.

The group average error for both methods was of the approximate magnitude of .3 semitone. This is greater than Bachem's estimate of .1 semitone for his *Ss*.

(Manuscript received October 15, 1950)

References

1. Bachem, A. Various types of absolute pitch. *J. acoust. Soc. Amer.*, 1937, **9**, 146–151.
2. Bachem, A. Genesis of absolute pitch. *J. acoust. Soc. Amer.*, 1939, **11**, 434–439.
3. Bachem, A. Chroma fixation at the ends of the musical frequency scale. *J. acoust. Soc. Amer.*, 1948, **20**, 704–705.
4. Bachem, A. Note on Neu's review of the literature on absolute pitch. *Psychol. Bull.*, 1948, **45**, 161–162.
5. Edwards, A. L. *Statistical analysis.* New York: Rinehart, 1946.
6. Neu, D. M. A critical review of the literature on "absolute pitch." *Psychol. Bull.*, 1947, **44**, 249–266.
7. Petran, L. A. Experimental study of pitch recognition. *Psychol. Monogr.*, 1932, **42**, No. 6.
8. Young, R. W., and Loomis, A. Theory of the chromatic stroboscope. *J. acoust. Soc. Amer.*, 1938, **10**, 112–118.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The issues listed below are available for sale

Year	Volume	Available Numbers						Price per Number	Price per Volume
1916	1	1	2	3	4	5	6	$1.25	$7.00
1917	2	1	2	–	4	5	–	$1.25	$5.00
1920	3	–	2	3	4	5	6	$1.25	$6.25
1921	4	1	2	3	4	5	6	$1.25	$7.00
1922	5	–	–	3	4	5	6	$1.25	$5.00
1923	6	–	–	–	–	–	–	——	——
1924	7	1	2	–	–	5	–	$1.25	$3.75
1925	8	1	2	3	4	–	6	$1.25	$6.25
1926	9	–	2	3	–	5	6	$1.25	$5.00
1927	10	–	–	3	4	5	6	$1.25	$5.00
1928	11	1	2	3	4	5	6	$1.25	$7.00
1929	12	1	2	3	4	5	6	$1.25	$7.00
1930	13	1	2	3	4	5	6	$1.25	$7.00
1931	14	1	2	3	4	5	6	$1.25	$7.00
1932	15	1	2	3	4	5	6	$1.25	$7.00
1933	16	1	2	3	4	5	6	$1.25	$7.00
1934	17	1	2	3	4	5	6	$1.25	$7.00
1935	18	1	2	3	4	5	6	$1.25	$7.00
1936	19	1	2	3	4	5	6	$1.25	$7.00
1937	20	1	2	3	4	5	6	$1.25	$7.00
	21	1	2	3	4	5	6	$1.25	$7.00
1938	22	1	2	3	4	5	6	$1.25	$7.00
	23	1	2	3	4	5	6	$1.25	$7.00
1939	24	1	2	3	4	5	6	$1.25	$7.00
	25	1	2	3	4	5	6	$1.25	$7.00
1940	26	1	2	3	4	5	6	$1.25	$7.00
	27	1	2	3	4	5	6	$1.25	$7.00
1941	28	1	2	3	4	5	6	$1.25	$7.00
	29	1	2	3	4	5	6	$1.25	$7.00
1942	30	1	2	3	4	5	6	$1.25	$7.00
	31	1	2	3	4	5	6	$1.25	$7.00
1943	32	1	2	3	4	5	6	$1.25	$7.00
	33	1	2	3	4	5	6	$1.25	$7.00
1944	34	1	2	3	4	5	6	$1.25	$7.00
1945	35	1	2	3	4	5	6	$1.25	$7.00
1946	36	1	2	3	4	5	6	$1.25	$7.00
1947	37	1	2	3	4	5	6	$1.25	$7 00
1948	38	1	2	3	4	5	6	$1.25	$7.00
1949	39	1	2	3	4	5	6	$1.25	$7.00
1950	40	1	2	3	4	5	6	$1.25	$7.00
1951	41	By subscription, $14.00 (2 vols.)						$1.25	——
	42								
		List price, Volumes 1 through 40							$260.25
		30% discount							78.08
		Net price, Volumes 1 through 40							$182.17

Information about the Journal of Experimental Psychology: as noted in the table, the journal was not published in 1918 and 1919; from 1937 to 1943 it was published at the rate of two volumes per year; from 1944 to 1950, in one volume per year. Beginning in 1951 there are two volumes a year.

Information about prices: when all numbers are available, the price per volume throughout the series is $7.00. For incomplete volumes, the price is $1.25 for each available number. For foreign postage, $.25 per volume should be added. The American Psychological Association gives the following discounts on orders for any one journal:

10% on orders of $ 50.00 and over
20% on orders of $100.00 and over
30% on orders of $150.00 and over

Address orders and subscriptions to

AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Avenue N.W. Washington 5, D. C.

EARLY PUBLICATION

IN

APA JOURNALS

The policy of accepting articles for immediate publication (providing the editor accepts the article and the author is willing to pay the entire cost of increasing the next available issue by enough pages to add his article to the normal content) is now standard practice for all APA journals except *Psychological Abstracts* and the *American Psychologist.*

The actual charge made to the author includes three items:

1. A basic charge of so much per page. This is the minimum amount that it costs to add an additional page to the journal. For 1951 these costs are:

	Per Page
Journal of Abnormal and Social Psychology	$17.00
Journal of Consulting Psychology	14.00
Journal of Comparative and Physiological Psychology	11.00
Journal of Consulting Psychology	14.00
Journal of Experimental Psychology	14.00
Psychological Bulletin	17.00
Psychological Monographs: General and Applied	15.00*
Psychological Review	14.00

* Since each Psychological Monograph is printed separately, the author of one handled on an early publication basis can be charged exactly the cost of printing. The figure of $15.00 is an approximate one; the actual figure will be higher for very short monographs and lower for very long ones. The cost will also vary depending upon the amount of special composition and the illustrations used.

These charges are based upon several factors:

(a) The greater number of words on a particular journal page, the higher the cost per page. Conversely, the fewer words printed on the page, the lower the cost per page.

(b) The more copies which must be printed, the higher the cost.

(c) The more expensive the printer, the higher the cost. Compared to the factors listed above, this is not an important difference in the charges made.

2. The full cost of any cuts or other illustrative material, of special composition for tables, and of author's changes in proof.

3. The full cost of any reprints ordered. (Authors of early-publication articles do not receive any free reprints.)

Classics Among

PSYCHOLOGICAL MONOGRAPHS

Thorndike, E. L. The Mental Life of the Monkey. 1899, #15. *$.50*

Carr, Harvey. Visual Illusion of Movement During Eye Closure. 1905, #31. *$1.25*

Watson, John B. Kinaesthetic and Organic Sensations: Their Role in the Reactions of the White Rat to the Maze. 1907, #33, *$1.00*

Shepherd, William T. Some Mental Processes of the Rhesus Monkey. 1910, #52. *$.75*

Franz, Shepard Ivory and **Lafora, Gonzalo R.** On the Functions of the Cerebrum: The Occipital Lobes. 1911, #56. *$1.25*

Fernberger, Samuel W. On the Relation of the Methods of Just Perceptible Differences and Constant Stimuli. 1912, #61. *$1.00*

Boring, Edwin G. Learning in Dementia Praecox. 1913, #63. *$1.00*

Langfeld, Herbert S. On the Psychophysiology of a Prolonged Fast. 1914, #71. *$.75*

Franz, Shepard Ivory. I. Symptomological Differences Associated with Similar Cerebral Lesions in the Insane. II. Variations in Distribution of the Motor Centers. 1915, #81. *$1.50*

Peckstein, Louis Augustus. Whole versus Part Methods in Motor Learning. 1917, #99. *$.75*

Kjerstad, Conrad L. The Form of the Learning Curves for Memory. 1919, #116. *$1.35*

Tolman, Edward C. Retroactive Inhibition as Affected by Conditions of Learning. 1918, #107. *$.75*

MANY OF THE EARLY MONOGRAPHS ARE OUT-OF-PRINT.
ONLY A LIMITED QUANTITY OF THE ABOVE
NUMBERS ARE AVAILABLE.

AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Ave. N.W., Washington 5, D. C.